September 30, 2014

The Honorable Tom Dempsey
President Pro Tem, Missouri State Senate
201 West Capitol Avenue
State Capitol Building, Room 326
Jefferson City, Missouri 65101

The Honorable Timothy Jones
Speaker, Missouri House of Representatives
201 West Capitol Avenue
State Capitol Building, Room 308
Jefferson City, Missouri 65101

Dear Senator Dempsey and Representative Jones:

Section 160.526, RSMo., requires that the Commissioner of Education inform the President Pro Tem of the Senate and the Speaker of the House about the procedures to implement the statewide assessment system, including a report related to the reliability and validity of the assessment instruments, at least six months prior to the implementation of the statewide assessment system. In compliance with that statute, I am pleased to provide the following information about updates to the Missouri Assessment Program for the 2014-2015 school year.

For the past several years, the Department of Elementary and Secondary Education has been developing updated high quality assessments in English language arts and mathematics. In compliance with HB1490, signed by Governor Nixon in July, these assessments will be implemented in the spring of 2015. As prescribed by the legislature, in order to ensure that all Missouri high school students graduate college and career ready, the Department of Elementary and Secondary Education is committed to implementing a reliable and valid assessment that will assess the knowledge, skills, and competencies called for in *The Show-Me Standards*. We have consulted with national experts and involved Missouri teachers in the development process.

The State Board of Education approved the following Missouri Assessment Program for 2014-2015.

Grade-Level Assessments
- All grade-level English language arts, mathematics, and science summative assessments will be delivered online.

- The English language arts and mathematics summative assessments for grades 3 through 8 will use a Computer Adaptive Test (CAT) blueprint. The mathematics summative assessment for grades 3 through 8 will consist of approximately 33 items. The grades 3 through 5 English language arts assessments will consist of approximately 43 items and Grades 6 through 8 English language arts will consist of approximately 44 items. Grades 5 and 8 will take summative English language arts and mathematics assessments that will also include a performance task in addition to the CAT portion of the assessments.
- A digital library of formative assessment resources will be available to Missouri educators at no cost to school districts and charter schools.
- English language arts and mathematics interim assessment resources will be available to all Missouri public schools.

MAP-Alternate (MAP-A) Grade-Level Assessment for students with significant cognitive disabilities
- English language arts and mathematics MAP-A assessments will be online, computer-adaptive assessments.
- An instructionally embedded assessment model will be used for the MAP-A English language arts and mathematics. The comprehensive assessment system is designed to support student learning and to more validly measure what students with the most significant cognitive disabilities know and can do.

End-of-Course (EOC) Assessments
- English language arts and mathematics EOC assessments have been updated to reflect the Missouri Learning Standards.
- A Physical Science EOC has been added to the Science subject area.

College and Career Readiness
- The administration of the ACT® Plus Writing to all public school 11th grade students at no cost in order to provide a statewide measure of college and career readiness.

Validity is the overarching component of the Missouri Assessment Program. The following excerpt is from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

*Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees.*

The validity and reliability of an assessment begins with the foundational content expectations, the development of items assessing the content, and the validation by educators that the items are accessible, fair, and representative of the content expectations. Additional evidence of validity and reliability is provided by item trials, pilots, and finally the official field testing of assessment items. The items are reviewed statistically for performance and fairness to all students. As the items move into operational administration, their performance is monitored to ensure that items continue to perform in a way that is fair to all students. The reporting and use of the data generated by assessments builds the final piece of the validity argument.

As an example, the following is an overview of the validation plan for the grade level assessment resources.

**The purposes of the *summative assessments* are to provide valid, reliable, and fair information about:**
- Students' English language arts and mathematics achievement with respect to those English language arts and mathematics Missouri Learning Standards measured by the English language arts and mathematics summative assessments.
- Whether students have demonstrated sufficient academic proficiency in English language arts and mathematics to be on track for achieving college and career readiness.
- Students' annual progress toward college and career readiness in English language arts and mathematics.
- How instruction can be improved at classroom, school, district, and state levels.
- Students' English language arts and mathematics proficiencies for federal and state accountability purposes and potentially for local accountability systems.
- Students' achievement in English language arts and mathematics across students and subgroups of students.

**The purposes of the *interim assessments* are to provide valid, reliable, and fair information about:**
- Student progress toward mastery of skills.
- Students' performance at the content cluster level, so that teachers and administrators can monitor student progress throughout the year and adjust instruction accordingly.
- Individual and group (e.g., school, district) performance at the claim level in English language arts and mathematics, to determine whether teaching and learning are on target.
- Student progress toward the mastery of skills measured in English language arts and mathematics across all students and subgroups of students.

**The purposes of the *formative assessment resources* are to provide measurement tools and resources to:**
- Improve teaching and learning.
- Monitor student progress throughout the school year.
- Help teachers and other educators align instruction, curricula, and assessment.
- Help teachers and other educators use the summative and interim assessments to improve instruction at the individual student and classroom levels.

We appreciate the support of the Missouri General Assembly in our endeavors and look forward to working with you in the future to raise the level of performance of students in Missouri's public schools.

If after reviewing the report you have any questions, please give us a call.

Sincerely,

Chris L. Nicastro
Commissioner of Education

Attachment

c:  State Board of Education

# Missouri Assessment Program
# Updates And Changes

2014-2015

**Table of Contents**

# I.
# Legislation and Policy Directives

## History of the Missouri Assessment Program

The Missouri Assessment Program (MAP) is designed to measure how well students acquire the skills and knowledge described in Missouri's Show-Me Standards – delineated in grade-levels, content areas, and courses within the Missouri Learning Standards (MLS) – in order for all high school graduates to be college and career ready. The assessments yield information on academic achievement at the student, class, school, district, and state levels. This information is used to diagnose individual student strengths and weaknesses in relation to instruction aligned to the MLS and to gauge the overall quality of education throughout Missouri.

The MAP traces its origin to the 1993 Outstanding Schools Act. This act required Missouri to create a statewide assessment system that measured challenging academic standards. Additionally, the State Board of Education directed the Missouri Department of Elementary and Secondary Education (the Department) to identify the knowledge, skills and competencies that Missouri students should acquire by the time they complete high school and to assess student progress toward those academic standards. The Department worked with teachers, school administrators, parents and business professionals from throughout the state to develop the "Show-Me Standards." From this act, grade-span assessments were created that measured Missouri's Show-Me standards. Originally, MAP was designed to be a grade-span test: Grades 3, 7, and 11 in Communication Arts, Grades 4, 8, and 10 in Mathematics, and Grades 3, 7, and 10 in Science.

In 2001, the federal No Child Left Behind (NCLB) legislation was enacted. In accordance with the NCLB legislation, student performance, reported in terms of proficiency categories, is used to determine the adequate yearly progress (AYP) of the school, district, and state using student performance results from the Missouri Assessment Program. NCLB required states to develop grade-level tests in both Reading and Mathematics to be administered in Grades 3 through 8 and once in high school. It also required that states have in place Science assessments to be administered at least once in Grades 3 through 5, once in Grades 6 through 9, and once in high school.

Beginning with the 2008-2009 school year, Missouri began administering End-of-Course (EOC) assessments in lieu of High School grade-level assessments. EOCs were implemented because they connected Missouri's high school accountability assessments directly to courses students take and they allowed school districts and charter schools to hold high school students accountable locally for their performance on the assessments. Algebra I, English II and Biology were the first EOCs administered. The following year, Government, American History, English I, Algebra II and Geometry became operational. The move to EOC assessments was also a move to online testing. In the first few years of EOC, districts had a choice between online testing and traditional paper/pencil assessments. EOCs moved fully online beginning in the fall of 2010.

The 2014-2015 school year is another time of transition for the Missouri Assessment Program. Grade-Level assessments in English language arts and mathematics in grades 3-8 and science in grades 5 and 8 will be administered fully online for the first time. In addition the English language arts and mathematics Grade-Level and EOC assessments will be aligned to the updated English language arts and mathematics Missouri Learning Standards.

**II.**
**Establishing a Foundation for the**
**Missouri Assessment Program**

**The Show-Me Standards**

The Show-Me Standards provide the basis for the MAP. These standards are designed to ensure that high school graduates can lead productive, fulfilling and successful lives as they continue their education, enter the work force and assume civic responsibilities. They set high expectations for learning and instruction, and encourage the development of challenging curricula in schools throughout the state. Beginning with the passage of the Outstanding Schools Act in 1993, hundreds of Missouri teachers, parents, and business professionals participated in the process of developing the Show-Me Standards, which were approved by the State Board of Education in January 1996.

There are a total of 73 Show-Me Standards (Appendix A). Forty of these are knowledge (content) standards, intended to delineate a solid foundation of knowledge and skills in the traditional subject areas (reading, writing, mathematics, world and American history, government, geography, science, health, physical education, and fine arts). The remaining 33 standards are process standards that require students to demonstrate and apply their content knowledge in a variety of situations. The process standards are grouped under four broad goals that are relevant to all content areas.

The Show-Me Standards were developed with the understanding that in order to be successful and productive, Missouri's students must have a solid foundation of knowledge and skills, as well as the ability to apply their learning to the kinds of problems and situations they will encounter after graduating from high school. The standards promote the concept that active, hands-on learning will benefit students of all ages. When basic knowledge and skills are integrated and applied in practical and challenging ways across subject areas, learning becomes engaging and motivating. Such learning stays in the mind long after tests are over and classrooms are left behind.

Although the Show-Me Standards define concepts that are significant to success in school, society and the workplace, they do not specify everything students should learn in school. The Show-Me Standards provide a solid foundation on which districts can build a challenging curriculum that will help all students reach their maximum potential.

**Missouri Learning Standards**
School districts must ensure that their curricula and instructional programs address the Show-Me Standards. The Missouri Learning Standards (MLS) were developed to provide guidance to school districts in this process. The Missouri Learning Standards represent an update to the Grade-Level Expectations (GLEs) and Course-Level Expectations (CLEs), which were developed in response to the move to grade-level assessments in mathematics and reading required by NCLB and implemented in the 2005-2006 school year. The GLEs and CLEs explicate the Show-Me Standards, providing specific targets for instruction and assessment. The MLS continue the precedence set by the GLEs and CLEs, by delineating the knowledge, skills, and abilities that students need to acquire at each grade level and/or course to progress toward the Show-Me Standards and college and career readiness along with targets for the statewide summative assessments. Missouri educators contributed to the development and review of the Missouri Learning Standards (MLS) to ensure that they reflect the realities of the classroom. The MLS provide the minimum content expectations for each grade and course andprovide the clarity and consistency teachers need to make sure their students are on track and equipped with the knowledge and skills they need for success.

The MLS do not tell teachers how to teach, but rather establish the minimum of what students need to learn. It is up to schools and teachers to decide how to best help students reach the standards. It is also important to note that the standards do not include everything that could or should be taught. Local districts are still designing their own curriculum and choosing which texts to read, along with a long list of other local education choices.

Additionally, the Missouri Learning Standards:

- Help colleges and professional development programs better prepare teachers.
- Establish a foundation for educators to work collaboratively with their peers to develop and share resources, expertise, curriculum tools and professional development.
- Guide educators toward curriculum and teaching strategies that will give students a deep understanding of the subjects and skills they need to learn.

The MLS take a spiraled approach, where key knowledge and concepts are introduced at early grade levels and then further developed in later grades. The content expectations encourage the integration of key concepts across subject areas. Important knowledge and skills traditionally taught in one subject area are reinforced and developed in other areas, broadening students' perspectives and understanding.

The Missouri Learning Standards are not a statewide mandated curriculum. Rather, they provide a framework for local curriculum development. The purpose of the MLS is to provide support for districts as they develop local curriculum guides that address the standards. Missouri law ensures local control of education. Each school district determines how its curriculum is structured. School districts have already developed curriculum guides reflecting the local approach to all students being college and career ready as envisioned by the Show-Me Standards.

**HB 1490 of 2014**
In accordance with HB 1490, which was passed in the 2014 legislative session and signed by Governor Nixon, advisory groups consisting of parents, classroom teachers and other education professionals, and career and technical education representatives are studying the learning standards and academic performance standards for the state.

Theses advisory work groups are reviewing the current learning standards in English language arts, mathematics, science, and history and government. For each subject area, there are two work groups; one for grades K-5 and another for grades 6-12. Members of the work groups were chosen by the Missouri Senate President Pro Tem and Speaker of the House, the Governor and Lieutenant Governor, the Commissioner of Higher Education, and the State Board of Education.

The work group will present suggested recommendations to learning standards to the State Board of Education in October 2015. Should the State Board of Education approve the learning standards recommendations, the updated learning standards will be implemented no earlier than the 2016-2017 school year. New assessments based on the standards will need to be developed.

**III.**
**About the Missouri Assessment Program**

**Subject Areas and Grade Levels for Statewide Assessment**

The Missouri Assessment Program (MAP) has statewide assessments that cover the following grades and content in the Show-Me Standards—mathematics, communication arts, science, and social studies.
- English language proficiency grades K-12 – ACCESS for ELLs
- Grades 3-8 English language arts and mathematics
- Grades 3-8, and 11 MAP-Alternative (MAP-A) English language arts and mathematics integrated yearlong assessment program for students with sever cognitive disabilities
- Grades 5 and 8 science
- Grades 5, 8, and 11 MAP-Alternative science for students with sever cognitive disabilities
- Algebra I, Geometry, Algebra II, English I, English II, American History, Government, Biology, Physical Science, and Personal Finance end-of-course assessments
- 11th grade ACT® Plus Writing

**Item Types**

Each state assessment may include up to four types of items: multiple-choice items, technology-enhanced items, constructed-response items and performance events/tasks.

**Multiple-choice items** present students with a question followed by four or five response options, one of which is correct. The advantages of these items are: 1) they are effective in measuring students' content knowledge and understanding; and, 2) a large number of these items can be administered and scored in a short amount of time, so that a wide range of knowledge and skills can be tested. The major limitation of multiple-choice items is that they do not adequately measure students' ability to apply what they know.

**Technology-enhanced items** are computer-delivered items that include specialized interactions for collecting response data. These include interactions and responses beyond traditional selected-response or constructed-response. The advantages of these items are: 1) they are effective in measuring students' content knowledge and understanding; 2) they allow students to demonstrate what they know in an authentic way; and, 3) they may be scored electronically.

**Constructed-response items** require students to supply (rather than select) an appropriate response. Students might be asked to provide a one-word answer, complete a sentence or show their work in solving a problem. In addition to measuring students' content knowledge, constructed-response items can provide some information about how students arrived at their answers. These items are more time consuming than multiple-choice items to administer and score; however, they provide more information about students' understanding and thinking.

**Performance events and performance tasks** measure students' knowledge, and their ability to apply that knowledge in problem situations. Performance events and tasks are one type of performance assessment. In its purest form, performance assessment requires students to demonstrate what they know. This type of assessment has been used for years in schools (e.g., band, business courses, and drivers' education).

The performance events and tasks used in the MAP may require a student to work through a complicated problem, or present a written argument. Depending on the demands of the task, students should be able to complete an event in 35 to 120 minutes depending on the grade, course, and or content area. Performance events and tasks often allow for more than one approach to get to a correct answer. The advantage of this type of assessment is that it provides insight into a student's ability to apply knowledge and understanding in various situations. The disadvantage is that performance events and tasks are often time consuming and costly to administer and score.

**Inclusion of Special Populations**

Throughout the development of the MAP, inclusion of special populations has been a goal. Accountability for all students was mandated through the Reauthorization of the Individuals with Disabilities Act (IDEA) 2004. Assessment development activities include educators with experience with disabilities, special education needs, and English language learners (ELL). Field tests in math and English language arts have included these students with special recruitment targets in order to ensure that special populations of students will be able to access the items and that the items are fair to all students.

The accommodations for the Grade-Level and End-of-Course assessments have been updated beginning with the fall 2014 EOC administration. With the move to online assessments for grades 3-8 many paper and pencil assessments accommodations are now part of the user interface. Online testing provides supports to students in three areas: Universal Tools, Designated Supports and Accommodations.

- Universal Tools are available to all students taking a Grade-Level or End-of-Course assessment. Examples of Universal Tools in the online administration platform include expandable passages, a highlighter, keyboard navigation, a mark for review function, a notepad, a protractor, a ruler, and spell check among other functionalities based on item type.
- Designated Supports are available to students when deemed appropriate by a team of educators. Examples of Designated Supports in the online administration platform include color contrast, color overly, magnification, masking, scribing, and turning off universal tools among other functionalities based on item type.
- Accommodations must appear in an IEP/504 plan and include American Sign Language, alternative response options (adapted keyboards, sticky keys, MouseKeys, etc.), multiplication tables, and speech-to-text among many other accommodations.

For Special Education students, the IEP team should choose all of the designated supports and accommodations a student will receive. Some designated supports and accommodations are only for ELL students.

The list of Grade-Level Accommodations, Supports and Tools may be found in Appendix B and the list of EOC Accommodations, Supports and Tools may be found in Appendix C.

ACT provides a variety of approved accommodations for students with IEPs and 504 plans. Assessments administered using ACT approved accommodations will result in college reportable ACT® scores. A student's IEP should reflect appropriate ACT-approved accommodations that will allow the student access to test content. High school ACT Test Accommodations Coordinators will receive training from ACT in fall 2014 to assist them in navigating the accommodations approval process.

Sample and Practice Assessment Items

The Department has traditionally provided practice and sample items to Missouri educators. This practice continues as the assessment program is updated for the 2014-2015 school year. The following table provides the type of sample and practice items available during the year to Missouri educators for professional development and use with students.

| | Released Items | Practice Items | Non-Secure Pre Tests | Non-Secure Interim Assessments |
|---|---|---|---|---|
| Grades 3-8 English language arts and mathematics | X | X | | X |
| Grades 3-8 MAP-A English language arts and mathematics | | X | | X |
| EOCs | | X | X | |
| ACT | | X | | |

Appendix D contains a sample of items from the Missouri Assessment Program.

**IV.**
**Missouri Assessment Program Test Development**

**Grades 3-8 English Language Arts and Mathematics Development**

A valid process of test development requires adherence to best practices, making possible the creation of tests that measure what is intended for a specified population. The various activities and analyses in the Test Development Phase can be viewed as establishing "Process Validity." Evidence that the assessments have met Process Validity requirements is found in the documentation of key development activities (listed more or less in order of sequence):

- Establishment of a validity framework, as outlined in our Research Agenda Recommendations located in Appendix E, to guide test development and ongoing research
- Development of a comprehensive set of content specifications based on the Missouri Learning Standards
- Development of test blueprints that specify number and types of questions to be presented to students
- Development of task models for items and stimuli that guide the writing and review of individual test items and passages
- Content and bias/sensitivity reviews, to ensure that items and passages are aligned to MLS content, are consistent with evidence statements of the content specifications, and are not biased in favor of or against students from different cultural and demographic backgrounds
- Cognitive lab on item types and accommodations to uncover issues and opportunities with modes of presentation, tools, and other item features. The final cognitive lab report may be found in Appendix F
- Establishment of a comprehensive accessibility framework and guidelines, found in Appendix G, to ensure that the assessments are accessible to the widest possible array of students
- Small-scale trials to investigate the use and feasibility of different item types. The final small-scale trials report may be found in Appendix H
- The procurement of software for computer adaptive test delivery, which is used to select items to be presented to students on the basis of both meeting the test blueprint and selecting items that maximize the accuracy of the student's score. The technical requirements for the system may be found in Appendix I
- Implementation of large-scale pilot tests in order to collect data on the initial performance of items and the testing platform software. The results of the pilot test may be found in Appendix J
- Implementation of large-scale field tests, described in Appendix K, to collect data on all items to evaluate their technical adequacy and their placement on a continuous growth scale from grades 3 through 11
- Analysis of the alignment, described in Appendix L, among all components of the assessment design, and ultimately between the Missouri Learning Standards and the tests students actually take
- Establishment of internal validity or the degree to which the test functions as required, has sufficient reliability, and sufficient ability to measure the intended content and not unintended content. Internal validity was investigated using Pilot Test results to determine whether or not a given content area test (ELA or mathematics) measured the intended construct and not unintended constructs. Essentially, this is an investigation as to whether or not the test is measuring primarily one construct (i.e., if it is uni-dimensional). As indicated in the dimensionality paper, included as Appendix M, the evidence strongly suggests that the ELA and

mathematics test are uni-dimensional. Test reliability will initially be modeled through simulations using the item pool after item review, which is due to be completed December 31, 2014. Operational test reliability will be reported in the technical manual following the first operational administration in spring 2015.

A total of one hundred and twenty-nine Missouri educators were selected to participate in the development of the 2014-2015 grade-level assessments. Eighteen Missouri educators were selected to serve as item writers along with fifty-two Missouri educators who selected to serve as reviewers. Eleven Missouri educators were selected to serve on content and bias review committees.

Two hundred and sixty-nine Missouri school districts and charter schools had the opportunity to participate in the spring 2013 item type pilot and two hundred and sixty-seven school districts participated in the spring 2014 item field test. The full list of participating school districts and charter schools that had the opportunity to participate in the spring 2013 pilot test may be found in Appendix N and the those school districts and charter schools that participated in the spring 2014 field test may be found in Appendix O.

Preliminary spring 2015 test designs and blueprints may be found in Appendix P.

## MAP-A English Language Arts and Mathematics Assessments

Missouri educators, especially those from the Missouri State Schools for the Severely Handicapped, have served as item writers, and as item and passage reviewers during the development of the new English language arts and mathematics MAP-A assessments used with the most significantly challenged students in the state. Using principles of evidence-centered design, item writers with expertise in English language arts, mathematics, and instruction for students with significant cognitive disabilities, developed testlets. Each testlet contains an engagement activity and three to seven questions.

Every testlet goes through multiple rounds of review by testing vendor assessment development staff, internal item reviewers, editors, and educators who serve as external reviewers. Each reviewing group is carefully trained to look for potential problems with the academic content, accessibility issues, and concerns about bias or sensitive topics. Staff review results from field tests to determine which testlets meet quality standards and are ready for operational assessment.

About 220 Missouri educators took part in the development of the 2014-2015 MAP-A assessments. Approximately 20 Missouri educators served as item writers along with about 177 Missouri educators serving as item reviewers.

192 school districts participated in the spring 2014 item field test. The full list of participating school districts, charter schools and school buildings may be found in Appendix Q.

Preliminary test designs and blueprints may be found in Appendix R.

## English Language Arts and Mathematics End-of-Course Assessments (EOCs)

In September 2007 and June 2008, Riverside Publishing conducted the first round of item-writing workshops to develop selected response (SR) items for English II and Algebra I as well as writing prompts for English II and PEs for Algebra I. These workshops were conducted at the Assessment Resource Center (ARC) in Columbia, Missouri and followed establish assessment industry best practices. Participants in

the workshops included Missouri educators, DESE staff, and Riverside Publishing TDSs. The workshops were held over a five-day period and were conducted with 15–20 teacher participants per content area. Teacher participants were selected by DESE to represent school districts throughout Missouri. Requirements to be an item writer included experience in classroom teaching and expert content knowledge. The content developed at the workshops was based on the Missouri Show-Me Standards and Course-Level Expectations.

The English II participants wrote selected response items associated with the passages that had been developed prior to the item-writing workshops. The Algebra I participants wrote SR items and PEs along with scoring guides.

During the item-writing workshops, Riverside Publishing Test Development Specialists (TDSs) conducted training sessions with the item writers and provided instructions on avoiding bias and stereotyping of groups and individuals on the basis of gender, race, ethnicity, religion, age, language, socioeconomic group, and disability. Riverside Publishing TDSs also trained item writers to write items that adhere to the principles of universal design, making the items accessible to the widest range of students. For example, items and passages were written using clear and concise language, and all art, graphs, and tables were labeled and were not overly crowded with extraneous information. Instruction was also provided on developing items at particular cognitive levels based on Norman Webb's Depth of Knowledge (DOK) levels. As items were produced, they were continuously reviewed, revised, edited, and evaluated by Riverside Publishing TDSs and DESE staff. Item writers who generated high-quality work on or ahead of schedule were given additional assignments.

All items and passages went through several rounds of internal reviews, including content and editorial reviews. Riverside Publishing TDSs reviewed each item with respect to alignment, clarity, and correspondence with item specifications.

Following item writing, twenty Missouri educators participated in a content and bias review process for each content area. The committee members read and reviewed each item. Discussions were held about whether the items met the criteria listed above. The committees then rejected or revised any items they deemed unsatisfactory. If there was disagreement about how to proceed with an item, the Riverside Publishing facilitator polled the group and followed the direction of the majority. Approximately 95% of the items were accepted (as–is or with edits) by the content and bias committees.

Similar workshops and reviews were subsequently held for the rest of the EOC assessments.

In 2012 and 2013 the English language arts and mathematics EOC item banks were aligned to the Missouri Learning Standards by panels of Missouri educators led by the Department's Directors of English Language Arts and Mathematics. The content directors then worked with the Department's testing vendor, CTB/McGraw-Hill's, test development specialists to design a test blueprint for the mathematics and English language arts EOCs similar to the previous assessments, but aligned to the Missouri Learning Standards. The test design and blueprints may be found in Appendix S.

**Physical Science EOC**

During the spring 2013 legislative session an increase in the assessment appropriation included money targeted for the development and launch of an additional science EOC. The Department partnered with the Iowa Testing Program located at the University of Iowa to develop a test blueprint and lease items to populate a Missouri unique EOC aligned to Missouri's physical science Missouri Learning Standards.

The assessment is an EOC that may be used at school district and charter school discretion. The test design and blueprints may be found in Appendix S.

**V.**
**Administering and Scoring the Missouri Assessment Program**

### Administering the State Assessment

The 2014-2015 Missouri Assessment Program brings additional resources to Missouri public and charter school educators. A digital library of formative resources and interim assessments for grade-level English language arts and mathematics will be available for educators to use as appropriate based on their local curriculum and instructional plan. These resources are provided at no charge to school districts and educators through the generosity of the Missouri Legislature's appropriation and the Governor's support. The interim assessments will be available in the same platform as the summative assessment to allow for complete educator and student comfort with the switch to online testing. The interim assessment data will be available only to school districts and charter schools.

Missouri school districts and charter schools will administer the MAP in designated grade levels, content areas, and courses as they have in the past. All assessments will be administered online. Testing windows may be found in Appendix T.

### Scoring the State Assessments

Several methods will be used to score the different components of state assessment. Multiple-choice and technology enhanced items will be machine-scored. Constructed-response items and performance events and performance task will be hand-scored by human readers. To ensure that the state assessments are scored quickly, and that the results are returned to districts in a reasonable amount of time, students' responses to the constructed-response items and performance events will be read by professional item scorers. This scoring will be organized and conducted by Missouri's assessment administration, scoring, and reporting vendors.

Hand-scoring the assessments is a <u>critical</u> piece in the development of the MAP. Much work is done months prior to the scoring activity. During this time, Missouri educators and Department staff, in partnership with Missouri's assessment vendors develop and review "score points" to ensure consistent grading of the papers.  The Department monitors the reliability and validity of this scoring by receiving frequent reports from partner vendors reporting interrater reliability.

There are several steps in the training and qualifying process to score Missouri's constructed response, performance events, and performance tasks. Potential scorers review the scoring guides and exemplar student responses. The potential scorers then work through several training rounds to accurately apply the scoring guide to specific types of responses. Once training is completed, the potential scorers must pass a qualifying round in order to score student papers.

Several technical methods are used to help maintain the accuracy of individual scorers during the scoring process. One method is the use of pre-scored papers which are periodically sent through the process without the scorers' knowledge. In the event that scorers do not score these "check sets" accurately, they are retrained, and monitored closely.

### Processing Data and Distributing Results

With the move to the online administration of grade-level assessments, Missouri public school elementary and middle school educators will experience what has become the norm for high school

educators in Missouri. State assessment results will be returned quickly after the close of the school district's or charter school's state testing window. The Department's testing vendor will return student English language arts and mathematics results to school districts within 10 business days of the close of the school district's testing window. The state will receive results in early July 2015 for the assessments from our partner vendors.

**VI.**
**Reporting and Using the Results of the**
**Missouri Assessment Program**

**Establishing Grade-Level English Language Arts and Mathematics Achievement Levels**

The grade-level English language arts and mathematics achievement levels will be established using a multi-step process designed to allow the participation of many Missouri educators. The process is designed to ensure that the resulting four achievement levels are valid, reliable, and fair measurements of college- and career-readiness for all students.

Achievement level setting will take place in three phases:

1. An online panel (scheduled for October 6-17) will allow up to 250,000 K-12 educators, higher education faculty, parents, and other interested parties to participate virtually in recommending achievement levels.
2. An in-person workshop (October 13–19) with panels of educators and other stakeholders working in grade-level teams will deliberate and make recommendations for the thresholds of the four achievement levels.
3. The vertical articulation committee, a subset of the in-person workshop, will then examine recommendations across all grades to consider the reasonableness of the system of cut scores.

The approach to achievement level setting emphasizes collaboration and transparency to establish a consistent means of measuring student progress on the interim and summative assessments. The online panel and the in-person workshop will provide an unprecedented opportunity to engage thousands of educators and interested stakeholders, raising awareness about the importance and rigor of the assessments. The results of the achievement level setting will be presented to the State Board of Education for their approval.

**Establishing Grade-Level English Language Arts and Mathematics Achievement Levels**

The a panel of educators with experience working with students with significant cognitive disabilities will convene in early June 2015, to establish achievement level cut points. Given the unique nature of the assessments and the student population, the achievement level setting teams for the English language arts and mathematics MAP-A assessments will be focus on addressing the following questions:

- What is the acceptable level of mastery certainty to be proficient on a node?
- When combining information across nodes, what threshold defines proficiency for the linkage level?
- How many linkage levels, across Essential Elements, must be mastered for each performance level?

The results of the achievement level setting will be presented to the State Board of Education for their approval.

**Validating 5th and 8th Grade Science Achievement Level Cut Points**

Given the change in the administration format of the 5th and 8th grade science assessments the Department will convene a panel of educators in early June 2015, to review the results of the first

statewide online administration of the 5<sup>th</sup> and 8<sup>th</sup> grade science assessments. The panelists will review the results and the application of the historical achievement level cut points, and will make a recommendation to the State Board of Education regarding any adjustments needed to cut points due to the administration of the assessments online.

**Validating English Language Arts and Mathematics Achievement Level Cut Points for EOCs**

Given the change in the blueprint of the English language arts and mathematics end-of-course assessments, the Department will convene a panel of educators in mid-February to review the results of the first statewide online administration of the updated English language arts and mathematics EOCs and the new physical science EOC. The panelists will review the results and the application of the historical achievement level cut points, and will make any necessary adjustments to the achievement level descriptors. In early June 2015, a small group of the panelist will reconvene to review the results of the spring administration of the updated and new EOCs, and make a recommendation to the State Board of Education regarding any adjustments needed to cut points due to the administration of the assessments online.

**Proposed Report Forms**

An individual student report (ISR) for parents will be delivered to public schools by the Department's testing vendor. The public schools will be responsible for distributing the individual student report to parents.   The ISR will describe the performance the student on the MAP. This report will provide the student's overall proficiency in a subject area, and how a student performed on a content area claim for each assessment taken. Students, parents, teachers and counselors can use the information included on this report, along with information gathered through local assessment programs, to improve a student's academic performance and to guide decisions about a student's educational options (e.g., which classes a student should take, in which school programs a student should participate). A sample student report may be found in Appendix U.

Classroom, grade, school, school district, and state results of the Missouri Assessment Program are reported through the Missouri Comprehensive Data System (MCDS). The MCDS is a resource provided by the Department that allows school personnel and the public to access education-related data.

The data made available to the public masks or hides data for groups with 10 or fewer students to protect confidential information about individual students, as required by federal law.

Three tools are available for data reviews are available within the MCDS portal:

- Quick Facts for basic reports and documents.
- Guided Inquiry for summary reports allowing simple filters.
- Advanced Inquiry for in-depth research and analysis.

The Missouri Comprehensive Data System may be found at http://mcds.dese.mo.gov.

**VII.**
**Technical Considerations**

## Validity of the MAP

In the process of test development, the term "validity" refers to the extent to which an assessment instrument measures what it is designed to measure. In order to ensure the validity of the MAP, the Department of Elementary and Secondary Education and our assessment partners are constructing the new English language arts and mathematics grade-level, MAP-A, and End-of-course assessments according to the highest standards of the industry. All development is occurring with the guidance of expert Technical Advisory Committees and the Department's technical and psychometric services vendor.

Validity of an assessment is enhanced when items are grade-appropriate. Field testing and substantial educator involvement in the development of the MAP items ensures that they are appropriate for the intended grade levels in Missouri. Following each field test, item statistics are generated to evaluate each item. Items are accepted, edited or, if necessary, discarded depending upon their performance on the field test.

Another factor which impacts the validity of an assessment instrument is item bias. Sound test development incorporates measures to eliminate any characteristics in an assessment that might unfairly influence student performance. A quality assessment must eliminate any influence by a student's cultural background, ethnicity, gender, race, or socio-economic status. All MAP items are reviewed for potential bias. The Department's assessment vendors produce supporting item statistics which indicate biases for or against particular subgroups in the student population.

Ultimately, accurate interpretation of test scores determines the overall validity of the assessment program. How well educators, parents, and the general public understand what the tests say is the "bottom line." Missouri educators, parents, and business professionals will be involved in defining and describing the levels of achievement that Missouri students are expected to attain.

## Reliability

The reliability of an assessment refers to the consistency of measurement it provides. Two types of reliability are being considered in the development of the MAP. The first is reliability across forms of the assessment. In other words, the assessment is reliable if a student would perform similarly on each of the three equivalent forms of a MAP subject area assessment. A common test blueprint is used to ensure that the difficulty and length of each form of the assessment are similar. Statistical equating procedures will be used to create reliable equivalent forms.

Because a portion of the MAP is performance based and must be hand-scored, inter-rater reliability is also being considered. Inter-rater reliability refers to the extent to which two different individuals would score a student's response in a similar manner. To accomplish high inter-rater reliability, concise scoring guides are created for each item and scorer training materials which provide clear examples of student work at each score point are selected. Each individual scoring student responses will be required to complete an extensive training session and pass a "qualifying round" of scoring. A variety of techniques will be used to maintain accuracy throughout the scoring of student responses, as well.

**VIII.**
**A Final Note**

In our fast-paced times, the general knowledge base and technology are changing and expanding at an amazing rate. Researchers are constantly identifying new and more effective educational methodologies. We must be responsive to these changes in order to provide the best possible opportunities for our children, the children of Generation Z. The Department believes that the adjustments to the 2014-2015 Missouri Assessment Program reflect best practices and industry standards. If Missouri hopes to provide the highest quality education for its students, then the state must continue to advocate for change that will promote educational progress.

Appendix A –    The Show-Me Standards

# The Show-Me Standards

Missouri students must build a solid foundation of factual knowledge and basic skills in the traditional content areas. The statements listed here represent such a foundation in reading, writing, mathematics, world and American history, forms of government, geography, science, health/physical education and the fine arts. This foundation of knowledge and skills should also be incorporated into courses in vocational education and practical arts. Students should acquire this knowledge base at various grade levels and through various courses of study. Each grade level and each course sequence should build on the knowledge base that students have previously acquired.

These concepts and areas of study are indeed significant to success in school and in the workplace. However, they are neither inclusive nor are they likely to remain the same over the years. We live in an age in which "knowledge" grows at an ever-increasing rate, and our expectations for students must keep up with that expanding knowledge base.

Combining what students must know and what they must be able to do may require teachers and districts to adapt their curriculum. To assist districts in this effort, teachers from across the state are developing curriculum frameworks in each of the content areas. These frameworks show how others might balance concepts and abilities for students at the elementary, middle and secondary levels. These models, however, are only resources. Missouri law assures local control of education. Each district has the authority to determine the content of its curriculum, how it will be organized and how it will be presented.

## Communication Arts

*In Communication Arts, students in Missouri public schools will acquire a solid foundation which includes knowledge of and proficiency in*

1. speaking and writing standard English (including grammar, usage, punctuation, spelling, capitalization)
2. reading and evaluating fiction, poetry and drama
3. reading and evaluating nonfiction works and material (such as biographies, newspapers, technical manuals)
4. writing formally (such as reports, narratives, essays) and informally (such as outlines, notes)
5. comprehending and evaluating the content and artistic aspects of oral and visual presentations (such as story-telling, debates, lectures, multi-media productions)
6. participating in formal and informal presentations and discussions of issues and ideas
7. identifying and evaluating relationships between language and culture

## Mathematics

*In Mathematics, students in Missouri public schools will acquire a solid foundation which includes knowledge of*

1. addition, subtraction, multiplication and division; other number sense, including numeration and estimation; and the application of these operations and concepts in the workplace and other situations
2. geometric and spatial sense involving measurement (including length, area, volume), trigonometry, and similarity and transformations of shapes
3. data analysis, probability and statistics
4. patterns and relationships within and among functions and algebraic, geometric and trigonometric concepts
5. mathematical systems (including real numbers, whole numbers, integers, fractions), geometry, and number theory (including primes, factors, multiples)
6. discrete mathematics (such as graph theory, counting techniques, matrices)

## Science

*In Science, students in Missouri public schools will acquire a solid foundation which includes knowledge of*

1. properties and principles of matter and energy
2. properties and principles of force and motion
3. characteristics and interactions of living organisms
4. changes in ecosystems and interactions of organisms with their environments
5. processes (such as plate movement, water cycle, air flow) and interactions of Earth's biosphere, atmosphere, lithosphere and hydrosphere
6. composition and structure of the universe and the motions of the objects within it
7. processes of scientific inquiry (such as formulating and testing hypotheses)
8. impact of science, technology and human activity on resources and the environment

**TURN OVER**

## Social Studies

*In Social Studies, students in Missouri public schools will acquire a solid foundation which includes knowledge of*

1. principles expressed in the documents shaping constitutional democracy in the United States
2. continuity and change in the history of Missouri, the United States and the world
3. principles and processes of governance systems
4. economic concepts (including productivity and the market system) and principles (including the laws of supply and demand)
5. the major elements of geographical study and analysis (such as location, place, movement, regions) and their relationships to changes in society and environment
6. relationships of the individual and groups to institutions and cultural traditions
7. the use of tools of social science inquiry (such as surveys, statistics, maps, documents)

## Fine Arts

*In Fine Arts, students in Missouri public schools will acquire a solid foundation which includes knowledge of*

1. process and techniques for the production, exhibition or performance of one or more of the visual or performed arts
2. the principles and elements of different art forms
3. the vocabulary to explain perceptions about and evaluations of works in dance, music, theater and visual arts
4. interrelationships of visual and performing arts and the relationships of the arts to other disciplines
5. visual and performing arts in historical and cultural contexts

## Health/Physical Education

*In Health/Physical Education, students in Missouri public schools will acquire a solid foundation which includes knowledge of*

1. structures of, functions of, and relationships among human body systems
2. principles and practices of physical and mental health (such as personal health habits, nutrition, stress management)
3. diseases and methods for prevention, treatment and control
4. principles of movement and physical fitness
5. methods used to assess health, reduce risk factors, and avoid high-risk behaviors (such as violence, tobacco, alcohol and other drug use)
6. consumer health issues (such as the effects of mass media and technologies on safety and health)
7. responses to emergency situations

# The Show-Me Standards

**Note to Readers:** What should high school graduates in Missouri know and be able to do? The Missourians who developed these standards wrestled with that question. In the end, they agreed that "knowing" and "doing" are actually two sides of the same coin. To perform well in school or on the job, one must have a good foundation of basic knowledge and skills. Equally important, though, is the ability to use and apply one's knowledge in real-life situations.

These standards (73 in all) are intended to define what students should learn by the time they graduate from high school. On this side are 33 "performance" standards, listed under four broad goals. On the reverse side are 40 "knowledge" standards, listed in six subject areas. Taken together, they are intended to establish higher expectations for students throughout the Show-Me State. These standards do not represent everything a student will or should learn. However, graduates who meet these standards should be well-prepared for further education, work and civic responsibilities.

**A**ll Missourians are eager to ensure that graduates of Missouri's public schools have the knowledge, skills and competencies essential to leading productive, fulfilling and successful lives as they continue their education, enter the workforce and assume their civic responsibilities. Schools need to establish high expectations that will challenge all students. To that end, the Outstanding Schools Act of 1993 called together master teachers, parents and policy-makers from around the state to create Missouri academic standards. These standards are the work of that group.

The standards are built around the belief that the success of Missouri's students depends on both a solid foundation of knowledge and skills *and* the ability of students to apply their knowledge and skills to the kinds of problems and decisions they will likely encounter after they graduate.

The academic standards incorporate and strongly promote the understanding that active, hands-on learning will benefit students of all ages. By integrating and applying basic knowledge and skills in practical and challenging ways across all disciplines, students experience learning that is more engaging and motivating. Such learning stays in the mind long after the tests are over and acts as a springboard to success beyond the classroom.

These standards for students are not a curriculum. Rather, the standards serve as a blueprint from which local school districts may write challenging curriculum to help all students achieve. Missouri law assures local control of education. Each school district will determine how its curriculum will be structured and the best methods to implement that curriculum in the classroom.

## GOAL 1

**Students in Missouri public schools will acquire the knowledge and skills to gather, analyze and apply information and ideas.**

*Students will demonstrate within and integrate across all content areas the ability to*

1. develop questions and ideas to initiate and refine research
2. conduct research to answer questions and evaluate information and ideas
3. design and conduct field and laboratory investigations to study nature and society
4. use technological tools and other resources to locate, select and organize information
5. comprehend and evaluate written, visual and oral presentations and works
6. discover and evaluate patterns and relationships in information, ideas and structures
7. evaluate the accuracy of information and the reliability of its sources
8. organize data, information and ideas into useful forms (including charts, graphs, outlines) for analysis or presentation
9. identify, analyze and compare the institutions, traditions and art forms of past and present societies
10. apply acquired information, ideas and skills to different contexts as students, workers, citizens and consumers

## GOAL 2

**Students in Missouri public schools will acquire the knowledge and skills to communicate effectively within and beyond the classroom.**

*Students will demonstrate within and integrate across all content areas the ability to*

1. plan and make written, oral and visual presentations for a variety of purposes and audiences
2. review and revise communications to improve accuracy and clarity
3. exchange information, questions and ideas while recognizing the perspectives of others
4. present perceptions and ideas regarding works of the arts, humanities and sciences
5. perform or produce works in the fine and practical arts
6. apply communication techniques to the job search and to the workplace
7. use technological tools to exchange information and ideas

## GOAL 3

**Students in Missouri public schools will acquire the knowledge and skills to recognize and solve problems.**

*Students will demonstrate within and integrate across all content areas the ability to*

1. identify problems and define their scope and elements
2. develop and apply strategies based on ways others have prevented or solved problems
3. develop and apply strategies based on one's own experience in preventing or solving problems
4. evaluate the processes used in recognizing and solving problems
5. reason inductively from a set of specific facts and deductively from general premises
6. examine problems and proposed solutions from multiple perspectives
7. evaluate the extent to which a strategy addresses the problem
8. assess costs, benefits and other consequences of proposed solutions

## GOAL 4

**Students in Missouri public schools will acquire the knowledge and skills to make decisions and act as responsible members of society.**

*Students will demonstrate within and integrate across all content areas the ability to*

1. explain reasoning and identify information used to support decisions
2. understand and apply the rights and responsibilities of citizenship in Missouri and the United States
3. analyze the duties and responsibilities of individuals in societies
4. recognize and practice honesty and integrity in academic work and in the workplace
5. develop, monitor and revise plans of action to meet deadlines and accomplish goals
6. identify tasks that require a coordinated effort and work with others to complete those tasks
7. identify and apply practices that preserve and enhance the safety and health of self and others
8. explore, prepare for and seek educational and job opportunities

TURN OVER

Appendix B –    Grade-Level Assessment Accommodations, Supports and Tools

## About The Updated Accommodations

The accommodations for the Grade-Level assessments have changed starting with the Spring 2015 Grade-Level administration.

What we previously knew as accommodations has now been split into three areas: Universal Tools, Designated Supports and Accommodations.
- Universal Tools are available to all students taking a Grade-Level or End-of-Course assessment.
- Designated Supports are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in an IEP/504 plan.

On the chart that follows, each tool, support and accommodation has a designation referring to the type of assessment it can be used for. Those designations are as follows:
- Online – If a tool, support or accommodation is designated online, it can only be used with the online assessment.
- Online (Not Embedded) – If a tool, support, or accommodation is designated online (not embedded), it can only be used with the online assessment but requires software not embedded in the system.
- Paper – If a tool, support, or accommodation is designated paper, it may only be used with the paper/pencil, Braille or large print assessments.
- Any – If a tool, support, or accommodation is designated any, it may be used with the online, paper/pencil, Braille or large print assessments.

For Special Education students, the IEP team should choose **all** of the designated supports and accommodations that a student will receive.

Some designated supports and accommodations are only for ELL students. ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

## Universal Tools

- The following is a list of universal tools for the Grade-Level and End-of-Course assessments.
- These tools are available to all students.

| Tool | Format | Description |
|---|---|---|
| **Break (Pause)** | Online | The system allows all students to pause the assessment for up to 20 minutes. There is no limit on the amount of times a student may use this tool. If the test is paused for more than 20 minutes the student will be prevented from returning to items already attempted. |
| | Any | All students may take breaks of up 20 minutes as needed. |
| **Calculator (For Calculator Allowed Items Only)** | Online | The system allows all students, on items where calculator use is allowed, to have access to an embedded digital calculator. |
| | Any | All students may have access, on items where calculator use is allowed, to a physical calculator. |
| **English Dictionary** | Online | The system allows all students access to an embedded English dictionary for use on the writing performance task. |
| | Any | All students may have access to a physical English dictionary for use on the writing performance task. |
| **Expandable Passages** | Online | The system allows all students to expand a passage or item so that it takes up a larger portion of the screen. |
| **Glossary (Grades 3-8 Math and ELA only)** | Online | The system allows all students to access an embedded glossary, which shows grade- and context-appropriate definitions of specific construct-irrelevant terms.<br>- This tool is not available for Grades 5 and 8 Science assessments. |
| **Highlighter** | Online | The system allows all students to have access to a highlighter for marking desired text, questions and answers. |
| | Paper | All students may have access to a physical highlighter. |
| **Keyboard Navigation** | Online | The system allows all students to navigate through the text by using the keyboard. |
| **Mark For Review** | Online | The system allows all students to mark an item for review. The flag is not saved if a student moves onto another segment or pauses the test for more than 20 minutes. |
| **Notepad (Scratch paper)** | Online | The system allows all students to use a digital notepad to make notes about an item. Notes are not saved when a student moves onto the next segment or pauses the test for longer than 20 minutes. During the writing performance task, notes are retained for all portions of the task. |
| | Paper | All students may have access to physical scratch paper to make notes about an item. Physical scratch paper should be collected and destroyed immediately upon the conclusion of the testing session. |
| **Protractor** | Online | The system allows all students to use an embedded protractor on specific items where appropriate. |
| | Paper | All students may have access to a physical protractor for use on specific items where appropriate. |
| **Ruler** | Online | The system allows all students to use an embedded ruler on specific items where appropriate. |
| | Paper | All students may have access to a physical ruler for use on specific items where appropriate. |
| **Spell Check** | Online | The system allows all students to use an embedded spell check feature on specific items where appropriate. The spell check feature only indicates that a word is misspelled; it does not provide the correct spelling. |
| **Strikethrough** | Online | The system allows all students to cross out answer options. |
| **Thesaurus** | Any | All students may have access to a physical thesaurus during the writing performance task. |

| | Universal Tools | |
|---|---|---|
| <div align="center">**Universal Tools**</div> | | |
| • The following is a list of universal tools for the Grade-Level and End-of-Course assessments. | | |
| • These tools are available to all students. | | |
| **Tool** | **Format** | **Description** |
| **Writing Tools** | Online | The system allows all students to use selected writing tools on specific items where appropriate. The tools include the ability to bold text, italicize text, create bullets points and an undo/redo feature. |
| **Zoom** | Online | The system allows all students to zoom in or zoom out on text or graphics to make they appear larger or smaller than the default size. |
| | Paper | All students may have access to devices that allow them to change the size of text, formulas, tables, graphics, etc. |

# Designated Supports

- The following is a list of designated supports for the Grade-Level and End-of-Course assessments.
- These supports are available to students when deemed appropriate by a team of educators.
- ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Support | Format | Description | Code |
|---|---|---|---|
| **Bilingual Dictionary** | Any | ELL students may have access to a physical bilingual dictionary for use on the writing performance task. | S431 |
| **Color Contrast** | Online | The system allows students to adjust background or font color based on student needs or preferences. | S101 |
| | Paper | Students may have the test presented to them printed in different colors based on student needs or preferences. | S102 |
| **Color Overlay** | Paper | Students may have a color transparency placed over the test presented to them based on student needs or preferences. | S103 |
| **Glossary (Grades 3-8 Math and ELA only)** | Paper | All students taking the paper based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment.<br>• This support is not available for Grades 5 and 8 Science assessments. | S104 |
| **Magnification** | Online - Not Embedded | The system allows students to use assistive technology devices to change the size of text, formulas, tables, graphics, etc. beyond the capabilities of the zoom tool. | S105 |
| **Masking** | Online | The system allows students to block off content that is not of immediate need or that may be distracting by using an embedded masking tool. | S106 |
| | Paper | Students may use a masking tool to block off content that is not of immediate need or that may be distracting. | S107 |
| **Read-Aloud (For all items in any subject, excluding ELA reading passages)** | Online | The system allows items in mathematics and English language arts to be read aloud to the student via embedded text-to-speech technology. The student can control the speed and volume of the voice. | S041 |
| | Online - Not Embedded | Students may use assistive technology text-to-speech software to allow all items in any subject, not including ELA reading passages, to be read aloud. | S042 |
| | Any | Students may have items in mathematics, science, and English language arts to be read aloud to them by a trained reader. Read Aloud of ELA reading passages requires an IEP or 504 plan. | S043 |
| | Any | ELL students may have items in mathematics, science, and English language arts to be read aloud to them in their native language by a trained translator. Read Aloud of ELA reading passages requires an IEP or 504 plan. | S111 |
| **Scribe (For all items in any subject, excluding ELA writing)** | Any | Students may dictate their responses to a trained scribe, who must follow the administration guidelines. Scribing of ELA writing requires an IEP or 504 plan. | S351 |
| **Separate Setting** | Any | Students may be allowed to test in a separate setting from other students. This includes testing individually or testing as part of a smaller group. | S501 |

| Designated Supports | | | |
|---|---|---|---|
| • The following is a list of designated supports for the Grade-Level and End-of-Course assessments. | | | |
| • These supports are available to students when deemed appropriate by a team of educators. | | | |
| • Designated supports must be turned on prior to testing. | | | |
| • ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2). | | | |

| Support | Format | Description | Code |
|---|---|---|---|
| **Translation** | Online | • The system allows ELL students to have the test directions for math translated through an embedded feature.<br>• The system allows ELL students to access translated glossaries for selected construct-irrelevant math items.<br>• The system allows ELL students to use stacked translations on selected construct-irrelevant math items. | S108 |
| | Any | • ELL Students may have test directions for math, science and social studies translated.<br>• ELL students may respond to any assessment in their native language. The responses must be translated and then transcribed by a trained scribe, who must follow the administration guidelines.<br>• ELL students taking the paper based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment. This glossary can be translated locally. | S109 |
| **Turn Off Universal Tools** | Online | The system allows test administrators to turn off universal tools that might be distracting to a student or that students are unable to use. | S100 |

- **Accommodations For Students With Disabilities**
- The following is a list of accommodations for the Grade-Level and End-of-Course assessments.
- The accommodation must appear in an IEP/504 plan to be allowed.
- ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Accommodation | Format | Description | Code |
|---|---|---|---|
| **Abacus** | Any | Students may have access to an abacus. | A391 |
| **Alternate Response Options** | Any | Students may respond to items using an alternate option, including but not limited to: Adapted Keyboards, StickyKeys, MouseKeys, FilterKeys, Adapted Mouse, Touch Screen, Head Wand, Switches. | A441 |
| **American Sign Language (ASL) (For math and science items and ELA listening items)** | Online | The system allows students to access math items and ELA listening items by viewing ASL video. | A051 |
| | Any | Students may have math, science, social studies items and ELA listening items translated into ASL. | A052 |
| **Braille** | Online | Students with visual impairments may read text via Braille. Refreshable Braille is available only for ELA. For math, Braille will be presented via embosser. ELA may be presented via embosser. | A011 |
| | Paper | Students with visual impairments may access the assessment via a Braille version. Tactile overlays and graphics tools may be used to assist the student in accessing the content. | A012 |
| **\*INVALIDATION\* Calculator GRADE 3 ONLY (For Non-Calculator Allowed Items Only) \*INVALIDATION\*** | Any | All students in Grade 3 may have access, on items where calculator use is not allowed, to a physical calculator. **NOTE: Use of this will result in invalidation – Student will receive lowest obtainable scale score (LOSS).** | A392 |
| **Calculator GRADES 4-8 ONLY (For Non-Calculator Allowed Items Only)** | Any | All students in Grades 4-8 may have access, on items where calculator use is not allowed, to a physical calculator. | A393 |
| **Closed captioning (ELA listening items)** | Online | The system allows students who are deaf or hard of hearing to access printed text that appears on the screen as audio materials are presented. | A101 |
| **Large Print** | Paper | Students with visual impairments may access the assessment via a Large Print version. | A021 |
| **\*INVALIDATION\* Multiplication Table GRADE 3 ONLY \*INVALIDATION\*** | Any | Students in Grade 3 may have access to a single digit multiplication table. **NOTE: Use of this will result in invalidation – Student will receive lowest obtainable scale score (LOSS)** | A394 |
| **Multiplication Table GRADES 4-8** | Any | Students in Grades 4-8 may have access to a single digit multiplication table. | A395 |
| **Paper Based Assessment** | Paper | Students may have access to a paper based version of the assessment. This can be accessed either by the complete assessment or by printing passages/stimuli/items on demand for the student as determined by the IEP/504. | A102 |

- **Accommodations For Students With Disabilities**
- The following is a list of accommodations for the Grade-Level and End-of-Course assessments.
- The accommodation must appear in an IEP/504 plan to be allowed.
- ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Accommodation | Format | Description | Code |
|---|---|---|---|
| **\*INVALIDATION\* Read-Aloud GRADES 3-5 ONLY (ELA reading passages) \*INVALIDATION\*** | Any | Students in grades 3-5 may have English language arts reading passages read aloud to them by a trained reader. **NOTE: Use of this will result in invalidation – Student will receive lowest obtainable scale score (LOSS).** | A041 |
| | Online - Not Embedded | Students in grades 3-5 may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud. **NOTE: Use of this will result in invalidation – Student will receive lowest obtainable scale score (LOSS).** | A042 |
| | Any | ELL students in grades 3-5 may have English language arts reading passages read aloud to them in their native language by a trained translator. **NOTE: Use of this will result in invalidation – Student will receive lowest obtainable scale score (LOSS).** | A111 |
| **Read-Aloud GRADES 6-8 and End-of-Course ONLY (ELA reading passages)** | Online | The system allows English language arts reading passages to be read aloud to the student via embedded text-to-speech technology. The student can control the speed and volume of the voice. | A043 |
| | Online - Not Embedded | Students may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud. | A044 |
| | Any | Students may have English language arts reading passages to be read aloud to them by a trained reader. | A045 |
| | Any | ELL students may have English language arts reading passages to be read aloud to them in their native language by a trained translator. | A112 |
| **Read-Aloud (ELA reading passages)** | Paper | Blind students in any grade who do not yet have adequate Braille skills may have ELA reading passages read aloud. | A046 |
| **Scribe (For ELA writing)** | Any | Students may dictate their responses to a trained scribe, who must follow the administration guidelines. | A351 |
| **Specialized Calculator (For Calculator Allowed Items Only)** | Any | Students may have access, on items where calculator use is allowed, to a specialized calculator, including talking calculators or Braille calculators, when appropriate. | A396 |
| **Speech-To-Text** | Online - Not Embedded | The system allows students to use voice recognition software so the student may use their voice to dictate responses or give commands. | A352 |

Appendix C –    End-of-Course Assessment Accommodations, Supports and Tools

## About The Updated Accommodations

The accommodations for the End-of-Course assessments have changed starting with the Fall 2014 EOC administration.

What we previously knew as accommodations has now been split into three areas: Universal Tools, Designated Supports and Accommodations.

- Universal Tools are available to all students taking a Grade-Level or End-of-Course assessment.
- Designated Supports are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in an IEP/504 plan.

On the chart that follows, each tool, support and accommodation has a designation referring to the type of assessment it can be used for. Those designations are as follows:

- Online – If a tool, support or accommodation is designated online, it can only be used with the online assessment.
- Online (Not Embedded) – If a tool, support, or accommodation is designated online (not embedded), it can only be used with the online assessment but requires software not embedded in the system.
- Paper – If a tool, support, or accommodation is designated paper, it may only be used with the paper/pencil, Braille or large print assessments.
- Any – If a tool, support, or accommodation is designated any, it may be used with the online, paper/pencil, Braille or large print assessments.

For Special Education students, the IEP team should choose **all** of the designated supports and accommodations that a student will receive.

Some designated supports and accommodations are only for ELL students. ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Universal Tools |||
|---|---|---|
| • The following is a list of universal tools for the Grade-Level and End-of-Course assessments. |||
| • These tools are available to all students. |||
| **Tool** | **Format** | **Description** |
| **Break (Pause)** | Any | All students may take breaks of up 20 minutes as needed. |
| **Calculator (For Calculator Allowed Items Only)** | Online | The system allows all students, on items where calculator use is allowed, to have access to an embedded digital calculator. |
| | Any | All students may have access, on items where calculator use is allowed, to a physical calculator. |
| **English Dictionary** | Any | All students may have access to a physical English dictionary for use on the writing performance task. |
| **Highlighter** | Online | The system allows all students to have access to a highlighter for marking desired text, questions and answers. |
| | Paper | All students may have access to a physical highlighter. |
| **Mark For Review** | Online | The system allows all students to mark an item for review. For End-of-Course assessments, flags are saved until the user indicates they are finished with the assessment. |
| **Notepad (Scratch paper)** | Online | The system allows all students to use a digital notepad to make notes about an item. |
| | Paper | All students may have access to physical scratch paper to make notes about an item. Physical scratch paper should be collected and destroyed immediately upon the conclusion of the testing session. |
| **Protractor** | Paper | All students may have access to a physical protractor for use on specific items where appropriate. |
| **Ruler** | Paper | All students may have access to a physical ruler for use on specific items where appropriate. |
| **Strikethrough** | Online | The system allows all students to cross out answer options. |
| **Thesaurus** | Any | All students may have access to a physical thesaurus during the writing performance task. |
| **Zoom** | Paper | All students may have access to devices that allow them to change the size of text, formulas, tables, graphics, etc. |

## Designated Supports

- The following is a list of designated supports for the Grade-Level and End-of-Course assessments.
- These supports are available to students when deemed appropriate by a team of educators.
- ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Support | Format | Description | Code |
|---|---|---|---|
| **Bilingual Dictionary** | Any | ELL students may have access to a physical bilingual dictionary for use on the writing performance task. | S431 |
| **Color Contrast** | Paper | Students may have the test presented to them printed in different colors based on student needs or preferences. | S102 |
| **Color Overlay** | Paper | Students may have a color transparency placed over the test presented to them based on student needs or preferences. | S103 |
| **Magnification** | Online - Not Embedded | The system allows students to use assistive technology devices to change the size of text, formulas, tables, graphics, etc. beyond the capabilities of a standard zoom tool. | S105 |
| **Masking** | Paper | Students may use a masking tool to block off content that is not of immediate need or that may be distracting. | S107 |
| **Read-Aloud** **(For all items in any subject, excluding ELA reading passages)** | Online - Not Embedded | Students may use assistive technology text-to-speech software to allow all items in any subject, not including ELA reading passages, to be read aloud. | S042 |
| | Any | Students may have items in mathematics, science, social studies and English language arts to be read aloud to them by a trained reader. Read Aloud of ELA reading passages requires an IEP or 504 plan. | S043 |
| | Any | ELL students may have items in mathematics, science, social studies and English language arts to be read aloud to them in their native language by a trained translator. Read Aloud of ELA reading passages requires an IEP or 504 plan. | S111 |
| **Scribe** **(For all items in any subject, excluding ELA writing)** | Any | Students may dictate their responses to a trained scribe, who must follow the administration guidelines. Scribing of ELA writing requires an IEP or 504 plan. | S351 |
| **Separate Setting** | Any | Students may be allowed to test in a separate setting from other students. This includes testing individually or testing as part of a smaller group. | S501 |
| **Translation** | Any | <ul><li>ELL Students may have test directions for math, science and social studies translated.</li><li>ELL students may respond to any assessment in their native language. The responses must be translated and then transcribed by a trained scribe, who must follow the administration guidelines.</li><li>ELL students taking the paper based, Braille or Large Print assessment may have access to a specific glossary, to be included with the assessment. This glossary can be translated locally.</li></ul> | S109 |

- **Accommodations For Students With Disabilities**
- The following is a list of accommodations for the Grade-Level and End-of-Course assessments.
- The accommodation must appear in an IEP/504 plan to be allowed.
- ELL students include those receiving services (RCV) or not receiving services (NRC). ELL students do not include those students in monitored status (MY1 or MY2).

| Accommodation | Format | Description | Code |
|---|---|---|---|
| **Abacus** | Any | Students may have access to an abacus. | A391 |
| **Alternate Response Options** | Any | Students may respond to items using an alternate option, including but not limited to: Adapted Keyboards, StickyKeys, MouseKeys, FilterKeys, Adapted Mouse, Touch Screen, Head Wand, Switches. | A441 |
| **American Sign Language (ASL) (For math, science, social studies items and ELA listening items)** | Any | Students may have math, science, social studies items and ELA listening items translated into ASL. | A052 |
| **Braille** | Paper | Students with visual impairments may access the assessment via a Braille version. Tactile overlays and graphics tools may be used to assist the student in accessing the content. | A012 |
| **Large Print** | Paper | Students with visual impairments may access the assessment via a Large Print version. | A021 |
| **Multiplication Table** | Any | Students taking End-of-Course assessments may have access to a single digit multiplication table. | A395 |
| **Paper Based Assessment** | Paper | Students may have access to a paper based version of the assessment. | A102 |
| **Read-Aloud (ELA reading passages)** | Online - Not Embedded | Students may use assistive technology text-to-speech software to allow ELA reading passages to be read aloud. | A044 |
| | Any | Students may have English language arts reading passages to be read aloud to them by a trained reader. | A045 |
| | Any | ELL students may have English language arts reading passages to be read aloud to them in their native language by a trained translator. | A112 |
| **Read-Aloud (ELA reading passages)** | Paper | Blind students in any grade who do not yet have adequate Braille skills may have ELA reading passages read aloud. | A046 |
| **Scribe (For ELA writing)** | Any | Students may dictate their responses to a trained scribe, who must follow the administration guidelines. | A351 |
| **Specialized Calculator (For Calculator Allowed Items Only)** | Any | Students may have access, on items where calculator use is allowed, to a specialized calculator, including talking calculators or Braille calculators, when appropriate. | A396 |
| **Speech-To-Text** | Online - Not Embedded | The system allows students to use voice recognition software so the student may use their voice to dictate responses or give commands. | A352 |

Appendix D –    Sample Assessment Items

For 1a-1b, select the symbol (<, >, or =) that should be placed in the box ☐ to make each statement true.

1a. $\left(372 \times \dfrac{5}{5}\right)$ ☐ $\left(372 \times \dfrac{8}{5}\right)$     ○ <     ○ >     ○ =

1b. $\left(372 \times \dfrac{5}{5}\right)$ ☐ $\left(372 \times \dfrac{1}{5}\right)$     ○ <     ○ >     ○ =

A carpenter used exactly 25 feet of wood to make 9 shelves of equal length. Each shelf measured between —

Ⓐ 1 and 2 feet.
Ⓑ 2 and 3 feet.
Ⓒ 3 and 4 feet.
Ⓓ 4 and 5 feet.

A survey was administered to 500 high school students to determine the type of music they prefer. The survey indicated that 22% prefer rock, 26% prefer hip hop, 29% prefer pop, and 23% selected "other." Which representation best illustrates the *number* of students preferring each type of music?

A) **Preferred Music**

| Type | Percent of Students |
|---|---|
| Rock | 22 |
| Hip Hop | 26 |
| Pop | 29 |
| Other | 23 |

C) **Preferred Music**

B) **Preferred Music**
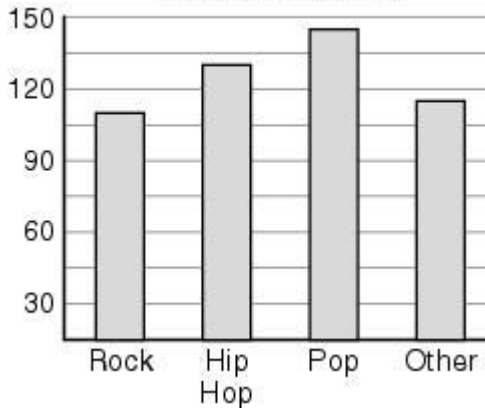
D) **Preferred Music**
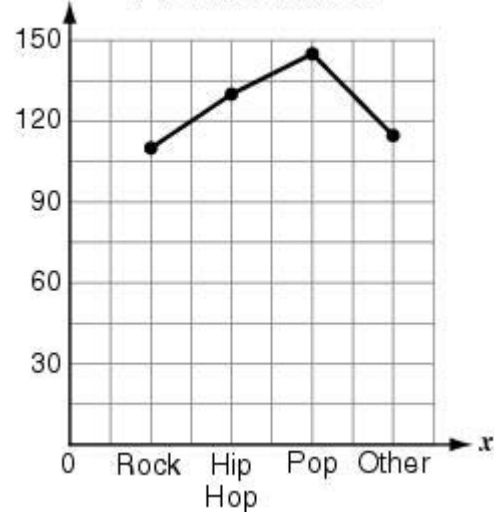
What is the value of the numerical expression below?

$$\sqrt{16} + \frac{24}{3} - 2^3$$

A) 4

B) 6

C) 8

D) 10

## What Are Coral Reefs?

The mention of coral reefs generally brings to mind warm climates, colorful fishes, and clear waters. However, the reef itself is actually a component of a larger ecosystem. The coral community is really a system that includes a collection of biological communities, representing one of the most diverse ecosystems in the world. For this reason, coral reefs often are referred to as the "rainforests of the oceans."

Corals themselves are tiny animals which belong to the group cnidaria (the "c" is silent). Other cnidarians include hydras, jellyfish, and sea anemones. Corals are sessile animals, meaning they are not mobile but stay fixed in one place. They feed by reaching out with tentacles to catch prey such as small fish and planktonic animals. Corals live in colonies consisting of many individuals, each of which is called polyp. They secrete a hard calcium carbonate skeleton, which serves as a uniform base or substrate for the colony. The skeleton also provides protection, as the polyps can contract into the structure if predators approach. It is these hard skeletal structures that build up coral reefs over time. The calcium carbonate is secreted at the base of the polyps, so the living coral colony occurs at the surface of the skeletal structure, completely covering it. Calcium carbonate is continuously deposited by the living colony, adding to the size of the structure. Growth of these structures varies greatly, depending on the species of coral and environmental conditions —ranging from 0.3 to 10 centimeters per year. Different species of coral build structures of various sizes and shapes ("brain corals," "fan corals," etc.), creating amazing diversity and complexity in the coral reef ecosystem. Various coral species tend to be segregated into characteristic zones on a reef, separated out by competition with other species and by environmental conditions.

Virtually all reef-dwelling corals have a symbiotic (mutually beneficial) relationship with algae called zooxanthellae. The plant-like algae live inside the coral polyps and perform photosynthesis, producing food which is shared with the coral. In exchange the coral provides the algae with protection and access to light, which is necessary for photosynthesis. The zooxanthellae also lend their color to their coral symbionts. Coral bleaching occurs when corals lose their zooxanthellae, exposing the white calcium carbonate skeletons of the coral colony. There are a number of stresses or environmental changes that may cause bleaching including disease, excess shade, increased levels of ultraviolet radiation, sedimentation, pollution, salinity changes, and increased temperatures.

Because the zooxanthellae depend on light for photosynthesis, reef-building corals are found in shallow, clear water where light can penetrate down to the coral polyps. Reef building coral communities also require tropical or sub-tropical temperatures, and exist globally in a band 30 degrees north to 30 degrees south of the equator. Reefs are generally classified in three types. Fringing reefs, the most common type, project seaward directly from the shores of islands or continents. Barrier reefs are platforms separated from the adjacent land by a bay or lagoon. Atolls rest on the tops of submerged volcanoes. They are usually circular or oval with a central lagoon. Parts of the atoll may emerge as islands.

Coral reefs provide habitats for a large variety of organisms. These organisms rely on corals as a source of food and shelter. Besides the corals themselves and their symbiotic algae, other creatures that call coral reefs home include various sponges; mollusks such as sea slugs, nudibranchs, oysters, and clams; crustaceans like crabs and shrimp; many kinds of sea worms; echinoderms like star fish and sea urchins; other cnidarians such as jellyfish and sea anemones; various types of fungi; sea turtles; and many species of fish.

*Item Prompt:*

Summarize the relationship between coral reefs and algae using details from the text.

# The Southland
excerpt from *White Fang*
## by Jack London

White Fang, written by Jack London, tells the story of a wild wolf dog's journey to domestication. When he is three years old, White Fang is found by Grey Beaver, a Native American living in Yukon Territory, Canada. White Fang pulls sleds to help Grey Beaver hunt and fish. Grey Beaver then sells White Fang to a new owner, who mistreats the wolf dog. Later, Weedon Scott becomes White Fang's owner and begins to further civilize the wolf dog by treating him with kindness.

White Fang landed from the steamer in San Francisco. He was appalled. Deep in him, below any reasoning process or act of consciousness, he had associated power with godhead. And never had the men seemed such marvelous gods as now, when he trod the slimy pavement of San Francisco. The log cabins he had known were replaced by towering buildings. The streets were crowded with perils—wagons, carts, automobiles; great, straining horses pulling huge trucks; and monstrous cable and electric cars hooting and clanging through the midst, screeching their insistent menace after the manner of the lynxes he had known in the northern woods.

All this was the manifestation of power. Through it all, behind it all, was man, governing and controlling, expressing himself, as of old, by his mastery over matter. It was colossal, stunning. White Fang was awed. Fear sat upon him. As in his cubhood he had been made to feel his smallness and puniness on the day he first came in from the Wild to the village of Grey Beaver, so now, in his full-grown stature and pride of strength, he was made to feel small and puny. And there were so many gods! He was made dizzy by the swarming of them. The thunder of the streets smote upon his ears. He was bewildered by the tremendous and endless rush and movement of things. As never before, he felt his dependence on the master, close at whose heels he followed, no matter what happened never losing sight of him.

But White Fang was to have no more than a nightmare vision of the city—an experience that was like a bad dream, unreal and terrible, that haunted him for long after in his dreams. He was put into a baggage-car by the master, chained in a corner in the midst of heaped trunks and valises. Here a squat and brawny god held sway, with much noise, hurling trunks and boxes about, dragging them in through the door and tossing them into the piles, or flinging them out of the door, smashing and crashing, to other gods who awaited them.

And here, in this inferno of luggage, was White Fang deserted by the master. Or at least White Fang thought he was deserted, until he smelled out the master's canvas clothes-bags alongside of him, and proceeded to guard them.

"'Bout time you come," growled the god of the car, an hour later, when Weedon Scott appeared at the door. "That dog of yourn won't let me lay a finger on your stuff."

White Fang emerged from the car. He was astonished. The nightmare city was gone. The car had been to him no more than a room in a house, and when he had entered it the city had been all around him. In the interval the city had disappeared. The roar of it no longer dinned upon his ears. Before him was smiling country, streaming with sunshine, lazy with quietude. But he had little time to marvel at the transformation. He accepted it as he accepted all the unaccountable doings and manifestations of the gods. It was their way.

There was a carriage waiting. A man and a woman approached the master. The woman's arms went out and clutched the master around the neck—a hostile act! The next moment Weedon Scott had torn loose from the embrace and closed with White Fang, who had become a snarling, raging demon.
"It's all right, mother," Scott was saying as he kept tight hold of White Fang and placated him. "He thought you were going to injure me, and he wouldn't stand for it. It's all right. It's all right. He'll learn soon enough."

And in the meantime I may be permitted to love my son when his dog is not around," she laughed, though she was pale and weak from the fright.

She looked at White Fang, who snarled and bristled and glared malevolently.

"He'll have to learn, and he shall, without postponement," Scott said.

He spoke softly to White Fang until he had quieted him, then his voice became firm.

"Down, sir! Down with you!"

This had been one of the things taught him by the master, and White Fang obeyed, though he lay down reluctantly and sullenly.

"Now, mother."
Scott opened his arms to her, but kept his eyes on White Fang.

"Down!" he warned. "Down!"


In paragraph 8, what does the word <u>placated</u> mean?

A. turned

B. scolded

C. soothed

D. distracted

Will Fish Farming Save Our Oceans?

   1 Only in the last few decades have people become aware that the ocean's **teeming** bounty is not, in fact, boundless. Until recently, almost all of the seafood eaten worldwide was harvested directly from the wild. People depended on the natural abundance and resilience of the oceans, rivers, and lakes. But as the human population has boomed, the need for fish as a food resource has also grown. For billions of people, fish are a primary source of protein. In some nations, such as the United States, where fish has traditionally made up only a small portion of the average diet, fish is seen as a healthier alternative to beef and pork. Around the world, the demand for seafood is on the rise.

   2 However, we can no longer rely on wild-caught seafood, as we have in the past. Overfishing, pollution, and loss of habitat have strained wild fish populations. There is now an urgent need for alternatives. One of these is aquaculture, or fish farming. But this solution is not without controversy.

What Is Aquaculture?

   3 Aquaculture means "farming or cultivating the water." The idea of farming fish is certainly not new. Like agriculture, it has been practiced since ancient times. But it was not until the 1960s and 1970s that aquaculture became a significant part of global production. It now accounts for more than 40 percent of the world's seafood.

   4 There are two basic types of aquaculture. The first is extensive aquaculture. Extensive aquaculturists set up their farms in oceans or bays, and natural currents keep the farm's water clean and full of oxygen. Oysters, mussels, and clams are raised this way, but so are some large finfish, such as salmon and tuna. How do the farmers prevent their mobile crops from escaping into the ocean?  The fish are kept in cages or "net pens" that are anchored to the ocean floor and can be densely stocked for higher production.

5 The other type of aquaculture is intensive. Freshwater fish such as catfish, tilapia, and carp are some of the species grown by intensive methods. This form of aquaculture relies on man-made ponds and advanced technology. One intensive fish farm in California grows 5 million pounds of tilapia per year in the middle of the desert! Enormous greenhouses with solar-heated tanks mimic the tilapia's natural environment. An advanced computer system removes waste, maintains temperature and oxygen levels, and feeds the fish on a regular schedule. An average-sized tilapia farm may have more than 200,000 fish in the tanks at any time.

A Solution . . .

6 Aquaculture seems to offer many advantages over traditional fishing. For one thing, fish farms might be able to reduce the pressure on wild fish populations. Also, some types of seafood are usually available in certain seasons only. Thanks to farms, these delicacies are available year-round. With careful breeding, farmers have produced "domesticated" fish that are fast growing and made-to-order. Now restaurants can plan menus knowing that fish of a certain kind and size will always be delivered. Reliable production has reduced the prices of many kinds of fish, making them more accessible as everyday food.

7 All this spells good news for the consumer. Aquaculture also seems to be good for developing nations. For example, on Zanzibar, an island off the eastern coast of Africa in the Indian Ocean, seaweed raised by aquaculture has become the leading export. Researchers are now developing techniques to add finfish and shellfish to this production. Local fish farms can provide more job opportunities and make cheaper seafood available to islanders and for export.

## . . . or Part of the Problem?

8 But aquaculture's supposed advantages may be too good to be true. In fact, fish farms may not be any healthier for the environment. The fish produced in farms must be fed. Their food is made from smaller species of "trash" fish, such as herring and anchovies, which are harvested directly from the ocean, further taxing wild fisheries. It takes two pounds of fish food to produce one pound of farmed fish–not a very economical ratio, to say the least!

9 Also, hundreds of thousands of fish are crammed together in these floating feedlots, as opponents call them. Fish farms create a lot of waste in the form of uneaten food, feces, dead fish, and chemicals. In extensive fish farming, this waste is flushed by the current into the surrounding ocean and bay, where it may affect the ecosystem in unknown ways. Intensive fish farmers often dump the waste from their artificial ponds and tanks into nearby waterways.

10 Fish farms not only affect the environment; they may also harm communities. The prospects for fish farming in the developing world seem promising. But the example of shrimp aquaculture in Southeast Asia casts doubts on its benefits for local residents. In Thailand and Vietnam, aquaculture has impaired rice farming, a traditional and far more efficient means of food production. Shrimp farms use up valuable fresh water and land resources vital to rice farmers, and the waste released into the environment has polluted water and farmland. Also, contrary to the hopeful claims of aquaculture advocates, the shrimp produced by these farms are not used to feed local populations cheaply. Instead, they are sold at high prices to the United States and other industrialized nations as luxury items.

## Future Outlook

11 Although there are compelling reasons to pursue aquaculture, it has created a whole new set of problems. With careful regulations and management, fish farms may eventually become the ideal solution to depleted fisheries, but there is much work to be done before this alternate source of seafood is truly sustainable.

Based on what you have read in the passage, which of these questions requires further evidence for support?

A)  What are the current methods of aquaculture?

B) What are the reasons for the reduction of fish in the wild?

C) What are some types of fish harvested through aquaculture?

D)  What are some ways to limit the negative effects of fish farming?

Read this sentence from the passage.

"Fish farms not only affect the environment; they may also harm communities."

Which question would best clarify the idea in the sentence?

A) How many fish can one fish farm produce in a single year?

B) What is the largest species of fish produced in the fish farms?

C) What are the long-term effects of waste products from fish farms on humans?

D) How do intensive aquaculture farms keep water conditions similar to oceans and lakes?

In a certain insect, round wings (R) are dominant to pointed wings (r). Which cross will produce the greatest number of genotypic and phenotypic variations?

A)  rr × rr

B) Rr × Rr

C) Rr × RR

D)  RR × RR

Polar bears swim across large expanses of ocean while hunting for seals, their main source of food. The bears use sea ice as resting spots during their long swims. However, the sea ice is rapidly melting as a result of global warming. Which statement describes what <u>most likely</u> will happen if global warming continues at its present rate?

A)  Polar bear and seal populations will both increase.

B) Polar bear populations will decrease, and seal populations will increase.

C) Polar bear populations will increase, and seal populations will decrease.

D)  Polar bear populations will decrease, and seal populations will remain the same.

Appendix E –     Grade-Level Research Agenda

# Smarter Balanced
# Assessment Consortium:
## Comprehensive Research Agenda

Report of Recommendations Prepared by Stephen G. Sireci

December 31, 2012

# Acknowledgments

# Table of Contents

# Smarter Balanced Assessment Consortium
# Comprehensive Research Agenda

## I. Introduction

In September 2010, the U.S. Department of Education awarded $175 million to the Smarter Balanced Assessment Consortium (Smarter Balanced) to develop assessments in English language arts (ELA) and mathematics that would "provide ongoing feedback to teachers during the course of the school year, measure annual student growth, and move beyond narrowly-focused bubble tests" (U.S. Department of Education, 2010). This award was part of the federal government's $4.35 billion Race to the Top competitive grant fund, which rewarded states for:

- Adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;

- Building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction;

- Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and

- Turning around our lowest-achieving schools. (U.S. Department of Education, 2009a, p. 2)

The goals of Smarter Balanced are comprehensive and are consistent with those of the Race to the Top Initiative. At the time of this report, Smarter Balanced represents a consortium of 25 states working together to develop cutting-edge ELA and mathematics assessments that feature computer-adaptive technology, technology-enhanced item formats, summative and interim assessments, and formative assessment resources. The assessment system being developed by the Consortium is designed to provide comprehensive information about student achievement that can be used to improve instruction and provide extensive professional development for teachers. The Smarter Balanced assessment system focuses on the need to strongly align curriculum, instruction, and assessment, in a way that provides valuable information to support educational accountability initiatives.

The specific goals of Smarter Balanced are described in its "Theory of Action," which is presented in Appendix A. The purpose of this report is to outline the research that should be conducted to (a) provide information to Smarter Balanced to help the Consortium accomplish its goals as it implements the program, and (b) evaluate the degree to which the Consortium is meeting its goals. Given that a large part of Smarter Balanced involves developing, administering, and scoring the assessments, and reporting the assessment results, much of the recommended research is based on the guidance provided by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), hereafter referred to as the Standards.

### Purposes of This Report

The purposes of this report are to inform Smarter Balanced of research that should be done to evaluate the degree to which the Consortium is accomplishing its goals and to demonstrate that the assessment system adheres to professional and federal guidelines for fair and high-quality assessment. The intent is to provide a comprehensive and detailed research agenda for the Consortium that includes suggestions and guidance for both short- and long-term research activities that will support Consortium goals.

To best inform the Consortium, we provide a description of the Standards, which were used as a framework for developing much of the research agenda. Integral to this description is a discussion of validity and the test validation process. We also reference the U.S. Department of Education's Standards and Assessments Peer Review Guidance (2009b), which stipulated the requirements for assessment programs to receive federal approval under the No Child Left Behind (NCLB) legislation. Although not described in this report, the research agenda also considered and is consistent with the Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (Yarbrough, Shulha, Hopson, & Caruthers, 2011) as well as the Guiding Principles for Evaluators (American Evaluation Association, 2004), which state that "evaluators aspire to construct and provide the best possible information that might bear on the value of whatever is being evaluated" (p. 1). The research agenda proposed here is designed to provide the best possible information to Smarter Balanced for understanding both the degree to which the Consortium is meeting its goals as well as what it can do to improve the system as it evolves.

In the remainder of this report, we (a) discuss the development of a validation plan that is consistent with the Standards and with the U.S. Department of Education's Standards and Assessments Peer Review Guidance; (b) list the primary purposes and goals of Smarter Balanced; (c) list the key validity issues associated with these purposes and goals; and (d) provide a description of studies that should be done to provide evidence regarding the degree to which Smarter Balanced assessments and activities are meeting the intended goals.

## II. Standards and Guidelines for Test Validation

### The Standards for Educational and Psychological Testing: A Validation Framework

There have been debates regarding what the term "validity" refers to, but for over 50 years three organizations—the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME)—have worked together to forge a consensus view of validity and provide guidance for developing and validating educational and psychological tests (Sireci, 2009). Currently, the Standards for Educational and Psychological Testing (AERA et al., 1999) define validity as "...the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 9). This definition emphasizes the importance of theory and empirical evidence to support the use of a test for a particular purpose. Thus, the research agenda for Smarter Balanced must be derived from the intended testing purposes and how assessment scores will be used.

The Standards describe the process of validation as that of developing a convincing argument, based on empirical evidence, that the interpretations and actions based on test scores are sound. Kane (1992, 2006) characterized this process as a validity argument, which is consistent with the validation process described by the Standards. For example,

> A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses . . . Ultimately, the validity of an intended interpretation . . . relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees . . . (AERA et al., 1999, p. 17)

This excerpt reinforces the Standards' emphasis that validation should center on test-score interpretation for specific uses. The research agenda developed for Smarter Balanced will be designed to fulfill the requirements of a sound validity argument as described by the Standards.

**The Standards' Five Sources of Validity Evidence.** To develop a sound validity argument, the Standards provide a validation framework based on five sources of validity evidence. These sources are validity evidence based on (a) test content, (b) response processes, (c) internal structure, (d) relations to other variables, and (e) consequences of testing.

Validity evidence based on *test content* refers to traditional forms of content validity evidence such as practice (job) analyses and subject-matter expert review and rating of test specifications and test items (Crocker, Miller, & Franks, 1989; Sireci, 1998), as well as newer "alignment" methods for educational tests that evaluate the links among curriculum frameworks, testing, and instruction (Bhola, Impara, & Buckendahl, 2003; Martone & Sireci, 2009). Evidence in this category is used to confirm that the tests that students take adequately represent the intended knowledge and skill areas. Confirming the degree to which the Smarter Balanced test specifications capture the intended Common Core State Standard (CCSS) and confirming that the items that students take adequately represent the areas delineated in the test specifications are examples of validity evidence based on test content that will be needed to build a strong validity argument for the Smarter Balanced assessments.

Validity evidence based on *response processes* refers to "evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by

examinees" (AERA et al., 1999, p. 12). Such evidence can include interviewing test takers about their responses to test questions, systematic observations of test response behavior, evaluation of the criteria used by judges when scoring performance tasks, analysis of item response time data, and evaluation of the reasoning processes that examinees use when solving test items (Embretson [Whitley], 1983; Messick, 1989; Mislevy, 2009). Such evidence will be needed to confirm that the Smarter Balanced assessments are measuring the cognitive skills that they intend to measure, and that students are using the targeted skills to respond to the test items.

Validity evidence based on *internal structure* refers to statistical analysis of item and sub-score data to investigate the primary and secondary (if any) dimensions measured by an assessment. Procedures for gathering such evidence include factor analysis (both exploratory and confirmatory) and multidimensional scaling. Internal structure evidence also evaluates the "strength" or "salience" of the major dimensions underlying an assessment, and so would also include indices of measurement precision, such as reliability estimates, decision accuracy and consistency estimates, generalizability coefficients, conditional and unconditional standard errors of measurement, and test information functions. In addition, analysis of differential item functioning (DIF), which is a preliminary statistical analysis to assess item bias, also falls under the internal structure category.

Evidence based on *relations to other variables* refers to traditional forms of criterion-related validity evidence, such as concurrent and predictive validity studies, as well as more comprehensive investigations of the relationships among test scores and other variables, such as multitrait-multimethod studies (Campbell & Fiske, 1959), and score differences across different groups of students, such as those who have taken different courses. These external variables can be used to evaluate hypothesized relationships between test scores and other measures of student achievement (e.g., test scores and teacher grades), to evaluate the degree to which different tests actually measure different skills, and the utility of test scores for predicting specific criteria (e.g., college grades). This type of evidence will be essential for supporting the validity of certain inferences based on scores from Smarter Balanced assessments (e.g., certifying college and career readiness).

Finally, evidence based on *consequences of testing* refers to evaluation of the intended and unintended consequences associated with a testing program. Examples of evidence based on testing consequences include investigations of adverse impact, evaluation of the effects of testing on instruction, and evaluation of the effects of testing on issues such as high school dropout and job applications. Other investigations of testing consequences relevant to the Smarter Balanced goals include analysis of students' opportunity to learn the CCSS, and analysis of changes in textbooks and classroom artifacts. With respect to educational tests, the Standards stress studying testing consequences. For example, they state,

> When educational testing programs are mandated . . . the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the use of the test, both intended and unintended, should also be examined by the test user. (AERA et al., 1999, p. 145).

Thus, it is important that validity evidence based on testing consequences is prominent in the Smarter Balanced research agenda.

**Using the Standards as a Validation Framework.** The Standards are considered to be "the most authoritative statement of professional consensus regarding the development and evaluation of educational and psychological tests" (Linn, 2006, p. 27). Therefore, they have great utility in

guiding a validity agenda. The validation research component of this comprehensive research agenda is based on crossing the intended purposes and use of Smarter Balanced assessments with the Standards' five sources of validity evidence. Therefore, the first step in determining the Smarter Balanced validity research agenda was to explicitly state its goals and purposes. These goals and purposes that are the focus of validation are described in Chapter III of this report.

### NCLB Peer Review Guidelines

One of the seven principles underlying the Smarter Balanced Theory of Action is the adherence "to established professional standards" (Smarter Balanced, 2010, p. 33). In addition to adhering to the Standards, the Consortium will also meet the requirements of the U.S. Department of Education's Peer Review process for NCLB assessments. Although these requirements are temporarily suspended as they undergo revision (Delisle, 2012), they remain important because they reflect the Department's most recent standards for ensuring quality and equity in statewide assessment programs. Thus, the research agenda incorporates much of the guidance provided in the Standards and Assessments Peer Review Guidance (U.S. Department of Education, 2009b). There is a great deal of overlap between the Standards and the U.S. Department of Education's Peer Review Guidance. However, the Guidance stipulates several important requirements that are highlighted in this research agenda. In particular, it requires:

- Providing evidence of the purpose of an assessment system and studies that support the validity of using results from the assessment system for their stated purpose and use (p. 42)

- Strong correlations of test and item scores with relevant measures of academic achievement, and weak correlations with irrelevant characteristics, such as demographics (p. 42)

- Investigations regarding whether the assessments produce intended or unintended consequences (p. 42)

- Documentation supporting evidence of the delineation of cut scores and the rationale and procedures for setting cut scores (pp. 21–22)

- Evidence of the precision of the cut scores & consistency of student classification (p. 44)

- Evidence of reliability for overall population and for each reported subpopulation (p. 44)

- Evidence of alignment over time through quality control reviews (p. 52)

- Evidence of comprehensive alignment and measurement of the full range of content standards and depth of knowledge and cognitive complexity (p. 54)

- Evidence that the assessment plan and test specifications describe how all content standards are assessed and how the domain is sampled to lead to valid inferences about student performance on the standards, individually and in the aggregate (using impartial experts in the process) (p. 54)

- Scores that reflect the full range of achievement standards (p. 57)

- Documentation to describe that the assessments are a "coherent" system across grades and subjects including studies establishing vertical scales (p. 34)

- Identification of how each assessment will provide information on the progress of students (p. 34)

The overlap of these requirements with the Standards is clear, and the anticipated revisions to this guidance will likely retain these key features. For example, in the recent letter informing states of the temporary suspension of peer review, the Department reiterated the following desired characteristics:

A high-quality assessment system [is] one that is "valid, reliable, and fair for its intended purposes; and measures student knowledge and skills against college- and career-ready standards in a way that

- Covers the full range of those standards, including standards against which student achievement has traditionally been difficult to measure;

- As appropriate, elicits complex student demonstrations or applications of knowledge and skills;

- Provides an accurate measure of student achievement across the full performance continuum, including for high- and low-achieving students;

- Provides an accurate measure of student growth over a full academic year or course; produces student achievement data and student growth data that can be used to determine whether individual students are college- and career-ready or on track to being college- and career-ready;

- Assesses all students, including English language learners and students with disabilities;

- Provides for alternate assessments based on grade-level academic achievement standards or alternate assessments based on alternate academic achievement standards for students with the most significant cognitive disabilities, consistent with 34 C.F.R. § 200.6(a)(2); and

- Produces data, including student achievement data and student growth data, that can be used to inform: determinations of school effectiveness for purposes of accountability under Title I; determinations of individual principal and teacher effectiveness for purposes of evaluation; determinations of principal and teacher professional development and support needs; and teaching, learning, and program improvement."

These characteristics of high-quality assessment systems were also considered in development of the comprehensive research agenda to ensure that evidence will be provided to demonstrate that the Smarter Balanced system meets these high standards.

## Other Validation Guidelines

In addition to the AERA et al. (1999) Standards and the U.S. Department of Education's (2009) Peer Review Guidance, there have been other seminal works that have influenced test validation practices. Messick's (1989) landmark chapter influenced the Standards and encouraged validators to focus on test use and the evaluation of testing consequences. Kane (1992, 2006), mentioned earlier, advanced Cronbach's (1988) notion of validation as an evaluation argument, and this notion is also embodied in the Standards. A recent addition to the validity literature is Bennett (2010), who expanded discussion of validation to include validation of a theory of action. This perspective is relevant to Smarter Balanced and is addressed in Chapter VIII. In short, this comprehensive research agenda incorporates many of the current theories and practices in test validation.

In addition to general guidelines on validation, there are also guidelines for specific testing applications. For example, the International Test Commission (ITC) produced Guidelines for

Translating and Adapting Tests (Hambleton, 2005; ITC, 2010), which are relevant to the evaluation of the Spanish-language versions of the Smarter Balanced mathematics assessments. There are also guidelines for universal test design (e.g., Johnstone, Altman, & Thurlow, 2006), and sensitivity review (e.g., Ramsey, 1993), which are relevant to the evaluation of the development of the Smarter Balanced assessments. Other documents consulted to guide this research agenda include Kane's (1994, 2001) criteria for evaluating standard setting studies (described further in Chapter IV) and the recent guidelines published by NCME (2012) on maintaining test integrity .

## III. Smarter Balanced Purpose Statements for Validation

As mentioned earlier, *validation* refers to gathering and evaluating evidence with respect to specific testing purposes. Thus, a first step in developing the comprehensive research agenda was identifying and articulating the intended purposes of Smarter Balanced. As the AERA et al. (1999) Standards state, "When educational testing programs are mandated by school, district, state, or other authorities, the ways in which test results are intended to be used should be clearly described . . ." (p. 168).

Although the Smarter Balanced Theory of Action described the overall goals of the Consortium, it was too general for evaluation or validation purposes. Thus, several steps were conducted to articulate the primary purposes and goals of Smarter Balanced that would be the focus of validation. These steps involved:

1. Extensive review of Smarter Balanced documentation;

2. Compiling a list of explicit claims, goals, and purposes;

3. Presenting this list to the Smarter Balanced Technical Advisory Committee (TAC);

4. Refining the list based on feedback;

5. Presenting the revised list to Smarter Balanced work groups;

6. Observing the Smarter Balanced Collaboration Conference and discussing goals, purposes, and validation plans with work groups, staff, and contractors;

7. Developing a draft list of Smarter Balanced goals and purposes to be the focus of validation;

8. Discussing this list with Smarter Balanced work groups via WebEx teleconferences; and

9. Revising the list based on work group input.

The identification of Smarter Balanced-specific goals began with the Theory of Action (Appendix A), but also involved a review of numerous Smarter Balanced documents, including the original Race to the Top application (Smarter Balanced, 2010), test specification documents (e.g., ETS, 2012a, 2012b), press releases, and requests for proposals (RFPs). More than 50 documents were reviewed in order to detect any stated claims, purposes, or goals. These reviews led to a preliminary list of goals and purposes that were presented to the Smarter Balanced TAC in July 2012. Feedback was received from the TAC and then from selected members of the Smarter Balanced Validation and Psychometrics/Test Design Work Group. Based on this feedback, refinements were made to the list of goals and purposes and were shared with Smarter Balanced leadership at the Collaboration Conference in September 2012. Further feedback was received, which included receipt of other documents that should be factored into the final articulation of goals and purposes.

Based on the observations and interaction with Consortium members, and the feedback provided by the TAC and the work group, a focus-group protocol was developed to involve Smarter Balanced leadership in the final articulation of testing purposes via WebEx teleconferences. Focus groups were held via WebEx in October 2012 with both the Validation and Psychometrics/Test Design Work Group and the Test Administration/Student Access Work Group. Excluding the facilitator, ten people participated in the first focus group (October 24, 2012) and sixteen people participated in the second (October 31, 2012). Each focus group was 90 minutes in duration. Following each focus group, draft purpose statements were sent to the participants via SurveyMonkey, and participants rated and commented on the appropriateness of the draft purpose statements. Based on these ratings and comments, the draft statements

were revised. These statements were presented to the TAC on December 12, 2012, and additional feedback was received and incorporated.

The final list of Smarter Balanced purpose statements that are the focus of validation follow. A description of the Smarter Balanced Theory of Action is presented in Appendix A to illustrate the degree to which the final list of purpose statements covers the major intentions stated in the Theory of Action.

The Smarter Balanced purpose statements for validation are separated into three categories that refer to (a) the summative assessments, (b) the interim assessments, and (c) formative assessment resources.

### The purposes of the Smarter Balanced *summative* assessments are to provide valid, reliable, and fair information about:

1. Students' ELA and mathematics achievement with respect to those CCSS measured by the ELA and mathematics summative assessments.

2. Whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA and mathematics to be on track for achieving college readiness.

3. Whether grade 11 students have sufficient academic proficiency in ELA and mathematics to be ready to take credit-bearing college courses.

4. Students' annual progress toward college and career readiness in ELA and mathematics.

5. How instruction can be improved at the classroom, school, district, and state levels.

6. Students' ELA and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems.

7. Students' achievement in ELA and mathematics that is equitable for *all students and subgroups of students*.

### The purposes of the Smarter Balanced *interim* assessments are to provide valid, reliable, and fair information about:

1. Student progress toward mastery of the skills measured in ELA and mathematics by the summative assessments.

2. Students' performance at the content cluster level, so that teachers and administrators can track student progress throughout the year and adjust instruction accordingly.

3. Individual and group (e.g., school, district) performance at the claim level in ELA and mathematics, to determine whether teaching and learning are on target.

4. Student progress toward the mastery of skills measured in ELA and mathematics *across all students and subgroups of students.*

### The purposes of the Smarter Balanced *formative assessment resources* are to provide measurement tools and resources to:

1. Improve teaching and learning.

2. Monitor student progress throughout the school year.

3. Help teachers and other educators align instruction, curricula, and assessment.

4. Help teachers and other educators use the summative and interim assessments to improve instruction at the individual student and classroom levels.

5. Illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning.

The remainder of this report centers on these purpose statements and their validation. The validation framework for the summative and interim assessments is based on the aforementioned five sources of validity evidence described in the Standards and involves crossing the purpose statements with each of the five sources. The formative assessment resources are not assessments per se, and so the research in support of their intended purposes extends beyond the five sources of validity evidence and follows a more traditional program evaluation approach.

As a prelude to Chapters V and VI, Tables 1 and 2 illustrate the validation framework for the Summative and Interim Assessments by crossing the purpose statements for each component with the five sources of validity evidence. The check marks in the cells indicate the type of evidence that is most important for validating each specific purpose. This presentation is extremely general, but indicates the comprehensiveness of the research agenda. It is also useful for understanding which sources of validity evidence are most important to specific purposes. For example, for purposes related to providing information about students' knowledge and skills, validity evidence based on test content will always be critical. For purposes related to classifying students into achievement categories such as "on track" or "college ready," validity evidence based on internal structure is needed, because that evidence includes information regarding decision consistency and accuracy.

Table 1. Validity Framework for Smarter Balanced *Summative Assessments*

| The purposes of the Smarter Balanced *summative* assessments are to provide valid, reliable, and fair information about: | Source of Validity Evidence | | | | |
|---|---|---|---|---|---|
| | Content | Internal Structure | Relations w/ Ext. Variables | Response Processes | Testing Consequences |
| 1. Students' ELA and mathematics achievement with respect to those CCSS measured by the ELA and mathematics summative assessments. | √ | √ | √ | √ | |
| 2. Whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA and mathematics to be on track for achieving college readiness. | √ | √ | √ | | √ |
| 3. Whether grade 11 students have sufficient academic proficiency in ELA and mathematics to be ready to take credit-bearing college courses. | √ | √ | √ | | √ |
| 4. Students' annual progress toward college and career readiness in ELA and mathematics. | √ | √ | √ | | √ |
| 5. How instruction can be improved at the classroom, school, district, and state levels. | √ | | | | √ |
| 6. Students' ELA and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems. | √ | √ | √ | | √ |
| 7. Students' achievement in ELA and mathematics that is equitable for *all students and subgroups of students*. | √ | √ | √ | √ | √ |

14

Table 2. Validity Framework for Smarter Balanced *Interim Assessments*

| The purposes of the Smarter Balanced *interim* assessments are to provide valid, reliable, and fair information about: | Source of Validity Evidence | | | | |
|---|---|---|---|---|---|
| | Content | Internal Structure | Relations w/ Ext. Variables | Response Processes | Testing Consequences |
| 1. Student progress toward mastery of the skills measured in ELA and mathematics by the summative assessments. | √ | √ | | √ | |
| 2. Students' performance at the content cluster level, so that teachers and administrators can track student progress throughout the year and adjust instruction accordingly. | √ | √ | | | √ |
| 3. Individual and group (e.g., school, district) performance at the claim level in ELA and mathematics, to determine whether teaching and learning are on target. | | √ | √ | | √ |
| 4. Student progress toward the mastery of skills measured in ELA and mathematics *across all students and subgroups of students.* | √ | √ | √ | √ | √ |

## IV. Essential Validity Elements for Summative and Interim Assessments

Before describing specific studies associated with each of the testing purposes listed in the previous chapter, it is important to first consider the fundamental validity information that is needed for *any* educational assessment program. These "essential elements" cut across the five sources of validity evidence and so deserve particular attention. The Standards describe such fundamental information as "evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees" (AERA et al., 1999, p. 17). Most of these essential elements fall under the categories of validity evidence based on test content (e.g., careful test construction) and internal structure (adequate score reliability, scaling, equating), but others, such as test administration and scoring, and careful attention to fairness, fall outside these two categories and do not neatly fit into the others. In addition to these fundamental elements, two other elements are essential: (a) equitable participation and access, and (b) test security.

In this chapter, we describe the types of information needed to confirm that these essential elements are adequately addressed in the research agenda. Because these elements refer to assessments, they are described in relation to the summative and interim assessments. However, "equal participation and access" is also important with respect to the formative assessment resources, which are discussed in Chapter VII.

In Table 3, we present a brief description of the validity evidence for the essential elements associated with the summative and interim assessments. Although the preceding quote from the Standards mentions adequate "reliability," we refer more generally to adequate "measurement precision" to underscore the need for measurement error to also be conceptualized in other frameworks such as item response theory (IRT) and generalizability theory.

The types of evidence listed in Table 3 will resurface when considering validity evidence for the specific purposes described earlier. This reoccurrence underscores the fundamental nature of these elements for supporting the use of Smarter Balanced assessments for their intended purposes. Most of these essential elements are typically addressed in technical manuals that support an assessment program. Descriptions of the types of studies to be conducted for each essential element follow.

### Careful Test Construction

As indicated in Table 3, validity evidence of careful test construction can come from a comprehensive audit of the test development process. This audit should be a comprehensive review of all test development activities, starting with the descriptions of testing purposes, operational definitions of the constructs measured, item development, content reviews, alignment studies, sensitivity reviews, pilot testing, item analyses, DIF analyses, item selection, item calibration, scoring rubrics for constructed-response items, and creation of test booklets (and clarity of test instruction). For adaptive assessments, the adequacy of the item selection algorithm, and the stopping rule, should also be reviewed.

Table 3. Validity Evidence Associated with Essential Elements for Summative and Interim Assessments

| Essential Element | Validation Evidence |
|---|---|
| Careful Test Construction | Audit of test development steps, including construct definition (test specifications and blueprints), item writing, content review, item analysis, alignment studies, and other content validity studies; review of technical documentation such as IRT calibration |
| Adequate Measurement Precision | Analysis of test information, conditional standard errors of measurement, decision accuracy, decision consistency, and reliability estimates for all reported scores |
| Appropriate Test Administration | Audit of test administration procedures, analysis of test irregularities, analysis of use and appropriate assignment of test accommodations |
| Appropriate Scoring | Audit of scoring procedures (hand, automated), inter-rater reliability analyses, rater drift (scale stability) analyses, computer/human comparisons (if relevant), generalizability studies, fairness for minorities |
| Accurate Scaling and Equating | Third-party verification of horizontal and vertical equating, IRT residual analysis, analysis of equating error, documentation of scaling and equating procedures, population invariance of equating |
| Appropriate Standard Setting | Comprehensive standard setting documentation, including procedural, internal, and external validity evidence for all achievement level standards set on assessments; includes criterion-related studies |
| Careful Attention to Fairness | Sensitivity review, DIF analyses, differential predictive validity analyses, comparability analyses (for language and disability accommodations), review of accommodation policies, implementation of accommodations, qualitative and statistical analyses of accommodated tests |
| Equitable Participation and Access | Analysis of participation rates, test accommodations, translations, and other policies |
| Adequate Test Security | Analysis of data integrity policies, test security procedures, monitoring of test administrations, analysis of cheating behavior, analysis of item exposure, review of chat rooms and websites for exposed items, review of anomalous results |

Examples of types of evidence that would be reviewed are presented in Table 4. Although a checklist format is used in Table 4, an audit would not simply check whether the activity was in place; rather, it would evaluate the quality of the activity.

Table 4. Sample Checklist for Audit of Test Construction Procedures

| Activity | Completed | Not Completed | Comments |
|---|---|---|---|
| Theory of Action/testing purposes clearly stated | | | |
| Development of test specifications sufficiently documented | | | |
| Item writers appropriately trained or recruited | | | |
| Items adhere to item writing guidelines | | | |
| Items reviewed for content quality and technical adequacy | | | |
| Content validity/alignment studies | | | |
| Sensitivity reviews | | | |
| Pilot study is adequate and representative | | | |
| Item analysis (classical) | | | |
| DIF analysis | | | |
| Item selection based on statistical and content criteria | | | |
| Item calibration | | | |
| Scoring rubrics for constructed-response items reviewed | | | |
| Adaptive item selection algorithm documented | | | |
| Test booklets are error-free | | | |

## Adequate Measurement Precision

*Measurement precision* extends the notion of reliability beyond a descriptive statistic for a test. It refers to the amount of expected variation in a test score, or classification based on a test score. Examples of this information include estimates of score reliability, standard errors of measurement, conditional standard errors of measurement, item and test information functions, conditional standard error functions, and estimates of decision accuracy and consistency. Estimates of score reliability include internal consistency estimates based on a single test administration (coefficient alpha, stratified alpha, marginal reliability), and those based on testing individuals more than once (test retest, parallel forms). The essential information needed for the Smarter Balanced assessments includes reliability estimates for all scores reported for students, estimates of decision consistency and accuracy for any reported achievement level results, and the traditional test information and standard error functions associated with IRT analyses. Generalizability studies that focus on specific sources of error will be important for identifying the sources of measurement error.

## Appropriate Test Administration

Evidence in this category involves review of test administration manuals and other aspects of the test administration processes. This review should include a review of the materials and processes associated with both standard and accommodated test administrations. Observations of test administrations, and a review of proctor and test irregularity reports, should also be included. The policies and procedures for granting and providing accommodations to students with disabilities and English language learners should also be reviewed, and case studies of accommodated test

administrations should be selected and reviewed to evaluate the degree to which the policies and procedures were followed.

## Appropriate Scoring

Validity evidence to confirm that the scoring of Smarter Balanced assessments is appropriate should include a review of scoring documentation. The Standards state that such documentation "should be presented . . . in sufficient detail and clarity to maximize the accuracy of scoring" (AERA et al., 1999, p. 47), as should the processes for selecting, training, and qualifying scorers. The scoring processes should also include monitoring of the frequency of scoring errors and how they are corrected. In terms of specific studies, evaluation of scorer reliability and score scale drift should be conducted. If any assessments are scored locally, the degree to which the scorers are trained, and the accuracy of their scores, should also be studied. Evidence in this category should also confirm that the routing of students during the adaptive exams is correct, and that all computerized scoring programs are accurate. The Standards *also* point out that one way to evaluate computerized scoring algorithms is to commission "an independent review of the algorithms by qualified professionals" (p. 70). Generalizability studies to locate sources of measurement error due to scoring will also provide important evidence.

## Accurate Scaling and Equating

Scaling and equating are essential activities for providing valid scores and score interpretations for Smarter Balanced assessments. Scaling activities include item calibration and creation of the standardized scale on which scores are reported. Equating activities will ensure that different forms of the assessments are on a common scale, as are scores reported over time. At the time of this writing, the summative assessments are intended to be vertically equated across grades. For the adaptive tests, the notion of a test "form" does not apply because the items are calibrated onto a common scale and can be assembled together uniquely for each examinee. This process requires that the items are correctly calibrated and that the IRT model sufficiently fits the data. Validity evidence for scaling and equating will include evaluation of the IRT model, confirming the hypothesized dimensionality of the assessments, evaluating equating documentation and estimates of equating error, evaluating the viability of a single construct (dimension) across grades, and, potentially, evaluating the invariance of the equating functions across important subgroups of students, such as students in different states. If funds are available, a "redundancy analysis," where an independent third party replicates the equating done by the contractor, would provide an important validity check on the accuracy of the equating.

## Appropriate Standard Setting

When achievement level standards are set on tests, test scores often become less important than the classifications that students receive. The standard setting literature is full of different methods for setting standards, but regardless of the method used, there must be sufficient validity evidence to support the classification of students into achievement levels. The Smarter Balanced summative assessments will use achievement levels, some of which will signify that students are "on track" to college readiness (grades 3–8) or "college ready" (grade 11). Kane (1994, 2001) wrote about gathering and documenting validity evidence for standards set on educational tests and categorized the evidence into three categories—procedural, internal, and external.

Procedural evidence for standard setting "focuses on the appropriateness of the procedures used and the quality of the implementation of these procedures" (Kane, 1994, p. 437). The selection of qualified standard setting panelists, appropriate training of panelists, clarity in defining the tasks and goals of the study, appropriate data collection procedures, and proper implementation of the method are all examples of procedural evidence.

Internal evidence for evaluating standard setting studies focuses on the expected consistency of results if the study were replicated. A primary criterion is the standard error of the cut score. However, calculation of this standard error is difficult due to dependence among panelists' ratings and practical factors (e.g., time and expense in conducting independent replications). Oftentimes evaluations of the variability across panelists within a single study, and the degree to which this variability decreases across subsequent rounds of the study, are presented as internal validity evidence. However, as Kane (2001) pointed out,

> A high level of consistency across participants is not to be expected and is not necessarily desirable; participants may have different opinions about performance standards. However, large discrepancies can undermine the process by generating unacceptably large standard errors in the cutscores and may indicate problems in the training of participants. (p. 73)

In addition to simply reporting the standard error of the cut score, Kane (2001) suggested that consistency can be evaluated across independent panels, subgroups of panelists, or assessment tasks (e.g., item formats), or by using generalizability theory to gauge the amount of variability in panelists' ratings attributed to these different factors. Another source of internal validity evidence proposed by Kane was to evaluate the performance of students near the cut score on specific items, to see if their performance was consistent with the panelists' predictions.

External validity evidence for standard setting involves studying the degree to which the classifications of students based on test scores are consistent with other measures of their achievement in the same subject area. External validity evidence includes classification consistency across different standard setting methods applied to the same test, tests of mean differences across examinees classified in different achievement levels on other measures of achievement, and the degree to which external ratings of student performance are congruent with the students' test-based achievement level classifications. It is likely that external validity evidence will be particularly important for validating the "college and career readiness" standards set on the summative assessments because several measures of college readiness already exist. In addition to classification consistency, the degree to which the constructs measured by these assessments overlap with the Smarter Balanced summative assessments, and the degree to which their definitions of readiness are similar, should be studied.

Some specific criteria that can be used to provide validity evidence for standard setting are summarized in Table 5. This table, adapted from Sireci, Hauger, Wells, Shea, & Zenisky (2009), illustrates the activities that should be conducted to (a) facilitate validity within the standard setting study, (b) evaluate the validity of the standard setting after it has been completed, or (c) do both.

Table 5. Summary of Criteria for Evaluating Standard Setting Studies

| Evidence | Criterion | Brief Explanation |
|---|---|---|
| Procedural | Care in selecting participants | Qualifications, competence, and representativeness of panelists; sufficient number of panelists |
| | Justification of standard setting method(s) | Degree to which methods used are logical, defensible, and congruent with testing purpose |
| | Panelist training | Degree to which panelists were properly oriented, prepared, and trained |
| | Clarity of goals/tasks | Degree to which standard setting purposes, goals, and tasks were clearly articulated |

| Evidence | Criterion | Brief Explanation |
|---|---|---|
| | Appropriate data collection | Data were gathered as intended |
| | Proper implementation | Method was implemented as intended |
| | Panelist confidence | Panelists understood tasks and had confidence in their ratings |
| | Sufficient documentation | Documentation of the entire process so that (a) it is understood and (b) it can be replicated |
| Internal | Sufficient inter-panelist consistency | Reasonable standard deviations and ranges of cut scores across panelists |
| | Decreasing variability across rounds | The variability across panelists' cut scores decreases across rounds—evidence of emerging consensus |
| | Small standard error of cut score (consistency within method) | Estimate of degree to which cut scores would change if study were replicated |
| | Consistency across independent panels | Estimate of degree to which cut scores would change if different panelists were used |
| | Consistency across panelist subgroups | Estimate of degree to which cut scores would change if specific types of panelists were used |
| | Consistency across item formats | Estimate of the consistency of cut scores across item formats (e.g., SR, CR items) |
| | Analysis of borderline students' performance on specific items | Degree to which expectations of hypothetical borderline students' performance are consistent with the performance of students near the cut scores |
| External | Consistency across standard setting methods | Degree to which results from different standard setting methods yield similar results |
| | Consistency across other student classification data | Degree to which classifications of students based on external data are congruent with classifications based on the cut scores |
| | Mean differences across proficiency groups on external criteria | Degree to which students classified into different achievement levels differ on other relevant variables |
| | Reasonableness | Degree to which cut scores produce results that are within a sensible range of expectations |

*Note: Adapted from Sireci et al. (2009).*

## Careful Attention to Fairness

Careful attention to fairness begins at the earliest stages of test development and includes many of the activities described in the previous section on careful test construction. One important aspect of

fairness is acknowledging the diversity within the student population when defining the constructs measured. Considerations of this diversity will reduce ethnocentricity in the construct definition and allow the development of accommodations policies that stay faithful to the construct measured. Sensitivity reviews and analysis of DIF and differential predictive validity are other important aspects of test fairness. Ensuring that students have the opportunity to learn material before it is tested and ensuring that a fair appeal process is in place are other important aspects of fairness. The presence of these practices and policies will be checked as part of the research agenda. The recent NCME document on data integrity underscores the need for testing programs to have policies and procedures to "ensure that all students have appropriate, fair, and equal opportunities to show their knowledge, skills, and abilities" (NCME, 2012, p. 3).

## Equitable Participation and Access

The Smarter Balanced system is designed for *all* students, and the intent is to provide flexibility and remove barriers that may inhibit students from taking the test and performing their best. The system is also designed to provide information widely, in transparent fashion, to all stakeholders. Equitable participation and access ensures that all students can take the test in a way that allows them to comprehend and respond appropriately.[1] The research agenda should include an analysis of participation rates across subgroups of students as well as a review of the procedures in place to ensure full participation. In particular, the degree to which Smarter Balanced offers sensible accommodations for students with disabilities and English language learners should be studied, as well as the availability and successful implementation of those accommodations. As stated in the recent NCME (2012) guidelines on test integrity, "Students who need accommodations due to language differences or students with disabilities may require appropriate modifications to materials and administrative procedures to ensure fair access to the assessment of their skills" (p. 3).

The U.S. Department of Education's Peer Review Guidance (2009b) provides additional guidance for confirming equitable participation and access. For example, it requires:

- Evidence of judgmental and data-based steps to ensure that assessments are fair and accessible to all students (p. 45)

- Evidence of how universal design or linguistic accommodations are incorporated (p. 45)

- Evidence that students with disabilities were included in the development process (p. 45)

- A policy on appropriate selection and use of accommodations (p. 47)

- Routine monitoring of accommodations used and ensuring that those used are used during instruction (p. 49)

- Checks of quality and consistency for accommodations given to English language learners (p. 49)

- Analysis of effect of usage of accommodations for English language learner students and students with 504s and IEPs (p. 49)

Another aspect of equitable participation and access is the provision of opportunities to retake an assessment. According to current policy, Smarter Balanced "will offer a retake opportunity on the CAT portion of the summative assessment for students who feel their scores are inaccurate or that believe the test was administered under non-standard circumstances" (Smarter Balanced, n.d.).

---

[1] Marty McCall, personal communication, December 22, 2012.

## Adequate Test Security

Test security is a prerequisite to validity. Threats to test security include cheating behaviors by students, teachers, or others who have access to testing materials. A lack of test security may result in the exposure of items before tests are administered, students copying or sharing their answers, or changing of students' answers to test questions. All of these behaviors have been observed in the past, and so those who value the validity of test scores worry about the prevalence of cheating behaviors. As described by NCME (2012), "When cheating occurs, the public loses confidence in the testing program and in the educational system, which may have serious educational, fiscal, and political consequences."

Thankfully, there are many proactive steps that testing agencies can take to reduce, eliminate, and evaluate cheating. The first step is to keep confidential test material secure and have solid procedures in place for maintaining the security of paper and electronic materials. The recent NCME (2012) document on data integrity outlined several important areas of test security. These areas include procedures that should be in place before, during, and after testing. The activities prior to testing include securing the development and delivery of test materials. Activities during testing include adequate proctoring to prevent cheating, imposters, and other threats. After testing, forensic analysis of students' responses and answer changes, and of aberrant score changes over time, are also beneficial. The goal of these security activities is to ensure that test data are "free from the effects of cheating and security breaches **and** represent the true achievement measures of students who are sufficiently and appropriately engaged in the test administration" (NCME, 2012, p. 3).

The evaluation of the test security procedures for the secure Smarter Balanced assessments will involve a review of the test security procedures and data forensics. The NCME (2012) document on test data integrity should be used to guide this evaluation. This document suggests that security policies should address:

> staff training and professional development, maintaining security of materials and other prevention activities, appropriate and inappropriate test preparation and test administration activities, data collection and forensic analyses, incident reporting, investigation, enforcement, and consequences. Further, the policy should document the staff authorized to respond to questions about the policy and outline the roles and responsibilities of individuals if a test security breach arises. The policy should also have a communication and remediation response plan in place (if, when, how, who) for contacting impacted parties, correcting the problem and communicating with media in a transparent manner. (p. 4)

With respect to specific studies that could evaluate security, in addition to an audit of test security policies, regular and systematic study of incorrect answer patterns for students who took the test in the same setting may be useful. However, with adaptive assessments, the probability of students receiving the same items at similar times is very low. Analyses of large score changes over time may be more useful, but it is important that any students, classes, or schools flagged for large score gains be considered innocent until proven guilty using external data (Wainer, 2011, chapter 8). Finally, given that most Smarter Balanced assessments will be delivered via computer, analysis of the time that students take to respond to items (e.g., are they correctly answering items in less time than it takes to read the item), and when tests are being accessed (are some tests accessed after hours?) will also provide important information regarding test security. Appendix C of the NCME (2012) document lists other examples of forensic analyses that could be conducted to evaluate test security.

## Summary of Essential Validity Elements

In considering the essential validity elements that are "relevant to the technical quality of a testing system" (AERA et al., 1999, p. 17), we arrive at many of the studies that should be contained within

the comprehensive research agenda. These studies will be highlighted again in the remaining chapters to underscore how they provide important information relevant to specific purposes of the Smarter Balanced Assessment Consortium, and are coordinated with the other studies described in the Introduction to this report.

# V. Validity Agenda for Summative Assessments

As described in Chapter III, there are seven purposes associated with the Smarter Balanced Summative Assessments that we recommend be the focus of validation. All of the studies discussed in Chapter IV that pertain to essential validity elements apply to these purposes. In this chapter, we relate these studies to each purpose statement and provide further descriptions where necessary.

It is important to note that each of the summative assessment purpose statements in Chapter III has the common preface "The purposes of the Smarter Balanced *summative* assessments are to provide valid, reliable, and fair information about . . ." In the sections that follow, we specify each purpose statement and then discuss the studies that should be done to provide the evidence to support the validity of the purpose. Within each purpose, the studies are organized by the Standards' five sources of validity evidence.

## Summative Assessment Purpose 1:

> Provide valid, reliable, and fair information about students' ELA and mathematics achievement with respect to those CCSS measured by the ELA and mathematics summative assessments.

As indicated in Table 1 (p. 14), validity evidence to support this purpose should come from at least three sources—test content, internal structure, and response processes. With respect to validity evidence based on test content, studies should be conducted to confirm that the content of the summative assessments adequately represents the CCSS intended to be measured in each grade and subject area. Appraisals of content domain representation and congruence to the CCSS must be made by carefully trained and *independent* subject-matter experts, not by employees of or consultants for the testing contractors. Validity evidence based on internal structure should involve analysis of item response data to confirm that the dimensionality of those data match the intended structure and support the scores that are reported. All measures of reliability, test information, and other aspects of measurement precision are also relevant. Validity evidence based on response processes should confirm that the items designed to measure higher-order cognitive skills are tapping into those targeted skills. The types of studies that are recommended for each of these three sources of validity evidence are described next.

**Validity Studies Based on Test Content.** Validity studies based on test content for the Smarter Balanced summative assessments need to evaluate the degree to which the assessments adequately measure the CCSS that they are designed to measure and in a way that conforms to the intended *evidence-centered design* (ECD; Mislevy & Riconscente, 2006). There should be at least two levels to the analysis. The first level would evaluate the degree to which the test specifications for the assessment sufficiently represent the intended CCSS. The second level of analysis should evaluate the degree to which the items administered to students adequately represent the test specifications. Studies relevant to these levels include traditional content validity studies (e.g., Crocker et al., 1989) and alignment studies (Bhola et al., 2003; Martone & Sireci, 2009; Porter & Smithson, 2002; Rothman, 2003; Webb, 2007). In Appendix B, we present brief descriptions of traditional content validity and alignment approaches and how they relate to one another.

*Evaluating test specifications.* To evaluate the appropriateness of the test specifications, the process by which the specifications were developed should be reviewed to ensure that all member states had input and that there was consensus regarding the degree to which the test specifications represent the CCSS targeted for the assessment. The degree to which states agree that the test specifications appropriately represent the CCSS, given the constraints of the assessment, could be ascertained by surveying curriculum specialists in the departments of education in the member states. Surveys could be constructed where these specialists would respond to selected- and open-response questions that would require them to comment on the degree to which the test specifications

adequately define the CCSS intended to be measured on the summative assessments, and the degree to which the relative weights of the cells in the test specifications reflect the corresponding emphases in the CCSS.

*Evaluating content and cognitive representation.* To evaluate the degree to which the summative assessments adequately represent the test specifications requires recruiting and training qualified and independent *subject-matter experts* (SMEs) in ELA, writing, and mathematics to review the CCSS within the test specifications and Smarter Balanced test items. At least two hypothesized aspects of the assessments need to be validated using SMEs. First is that the items are appropriately measuring the CCSS that they are designed to measure. Second is that the items are measuring the breadth of higher- and lower-order cognitive skills that they are designed to measure. There are a variety of methods that could be used to evaluate these aspects of content validity—some based on traditional notions of content validity, and others based on alignment methodology (Martone & Sireci, 2009). What the specific method is called is not important. What is important is that the tasks presented to the SMEs allow them to provide the data needed to evaluate the degree to which the assessments sufficiently represent the intended CCSS and the cognitive skills targeted by these standards.

To evaluate the degree to which each test item adequately represents (i.e., is aligned with) its corresponding CCSS, there are several studies that could be conducted, ranging from simply having SMEs match test items to claim areas (similar to Webb's categorical concurrence or Achieve's [2006] blueprint confirmation) to having the SMEs use a Likert-type rating scale to rate the congruence between each item and the CCSS that it is designed to measure. An example of the "matching" approach is presented in Figure 1, and an example of how the data from such a study could be summarized is presented in Figure 2. An example of the rating approach is presented in Figure 3; an example of how the rating scale data can be summarized is presented in Figure 4.

Regardless of the method chosen, appropriately summarizing the results of these content-based validity studies is important. Results should be analyzed at the item level to screen out or revise any items that have poor alignment ratings. More important, however, is aggregating the data so that the representation of the claims or assessment targets within each subject area can be evaluated.

In addition to the descriptive summaries of alignment, these studies should also compute congruence/alignment statistics. Such statistical summaries range from purely descriptive to those that involve statistical tests. On the descriptive end, Popham (1992) suggested a criterion of 7 of 10 SMEs rating an item congruent with its standard to confirm the fit of an item to its standard. This 70% criterion could be applied to the claim level and other aggregations of items. On the statistical end, several statistics have been proposed for evaluating item-standard congruence, such as Hambleton's (1980) item-objective congruence index and Aiken's (1980) content validity index. In addition, Penfield and Miller (2004) established confidence intervals for SMEs' mean ratings of content congruence.

Figure 1. Sample Item/Assessment Target Rating Form for Summative Assessment: Reading (Literary)

| Item # | Assessment Target (choose one for each item) | | | | | | |
|--------|-------------|---------------|------------------|------------------------|-------------------------|------------------------------|--------------|
| | Key Details | Central Ideas | Word Meanings | Reasoning & Evaluation | Analysis w/in, across Texts | Text Structures & Features | Language Use |
| 432 | | | | | | | |
| 433 | | | | | | | |
| 434 | | | | | | | |
| 443 | | | | | | | |
| 563 | | | | | | | |
| 578 | | | | | | | |
| 579 | | | | | | | |
| 580 | | | | | | | |
| 581 | | | | | | | |

From the matching approach (Figure 1), we can see how these data can inform us about the degree to which the assessment targets are represented by the items in a general sense. For example, in Figure 2, we see that the items associated with the assessment target "Analysis within and across Texts" were generally considered congruent with this target by the SMEs, but the items measuring "Language Use" were less congruent. Specific items could be revised or deleted to improve the representation of an assessment target. However, the matching approach does not give us information about *how well* the items measure their associated achievement target. Therefore, the rating scale approach is preferable, even though it may take slightly longer for the SMEs to provide those ratings.

Figure 2. Example Summary of Item/Assessment Target Congruence

| Assessment Target | # of Items | % of Items Classified Correctly by All SMEs | % of Items Classified Correctly by at Least 7 SMEs |
|-------------------|-----------|---------------------------------------------|----------------------------------------------------|
| Key Details | 22 | 45% | 86% |
| Central Ideas | 17 | 88% | 94% |
| Word Meanings | 33 | 55% | 97% |
| Reasoning & Evaluation | 25 | 48% | 80% |
| Analysis w/in, across Texts | 12 | 92% | 100% |
| Text Structures & Features | 21 | 71% | 90% |
| Language Use | 17 | 41% | 76% |
| **Average:** | | **56%** | **89%** |

Using the rating scale approach (Figure 3), we can get an idea of how well specific items, and the group of items comprising a content category or other level of the test specifications, adequately measure the intended standard or area, with respect to the characteristics of the rating scale. For example, the fictitious results in Figure 4 may suggest that the content categories have good representation with respect to the degree to which the items are measuring the CCSS within each

area. However, some specific items should be flagged for review and possibly revised or deleted. A similar rating task could be used to evaluate how well the items are measuring the intended cognitive skills. A cognitive skill dimension was not noted in the current test blueprints for the Smarter Balanced summative assessments, and so a cognitive skill classification such as that used in the Webb (1999), Achieve (2006), or Porter & Smithson (2002) alignment approaches could be adopted and arranged as a rating task, such as those presented in Figure 1 and Figure 3.

Figure 3. Example of SME Rating Task Assessing Item/CCSS Congruence

Directions: Please read each item and its associated benchmark. Rate how well the item measures its benchmark, using the rating scale provided. Be sure to circle one rating for each item.

| Item | Common Core State Standard (Grade 4 ELA) | How well does the item measure its CCSS? (circle one) | | | | | | Comments (Optional) |
|---|---|---|---|---|---|---|---|---|
| | | 1 (Not at all) | 2 | 3 | 4 | 5 | 6 (Very well) | |
| 226 | Refer to details and examples in a text when explaining what the text says explicitly and when drawing inferences from the text. | 1 | 2 | 3 | 4 | 5 | 6 | |
| 238 | Determine a theme of a story, drama, or poem from details in the text; summarize the text. | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1006 | Describe in depth a character, setting, or event in a story or drama, drawing on specific details in the text (e.g., a character's thoughts, words, or actions). | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1064 | Determine the meaning of words and phrases as they are used in a text, including those that allude to significant characters found in mythology (e.g., Herculean). | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1428 | Explain major differences between poems, drama, and prose, and refer to the structural elements of poems (e.g., verse, rhythm, meter) and drama (e.g., casts of characters, settings, descriptions, dialogue, stage directions) when writing or speaking about a text. | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1614 | Determine a theme of a story, drama, or poem from details in the text; summarize the text. | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1658 | Determine the meaning of words and phrases as they are used in a text, including those that allude to significant characters found in mythology (e.g., Herculean). | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1676 | Compare and contrast the point of view from which different stories are narrated, including the difference between first- and third-person narrations. | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1733 | Refer to details and examples in a text when explaining what the text says explicitly and when drawing inferences from the text. | 1 | 2 | 3 | 4 | 5 | 6 | |

Figure 4. Example Summary of Results from Item/CCSS Congruence Study

| Item | Content Category | Mean | Median | Aiken Index |
|------|------------------|------|--------|-------------|
| 226 | Reading-Literary | 4.2 | 4.0 | .89* |
| 238 | Reading-Literary | 5.3 | 5.0 | .91* |
| 1006 | Reading-Literary | 4.1 | 4.5 | .90* |
| 1064 | Reading-Literary | 3.5 | 4.0 | .91* |
| 1121 | Reading-Literary | 4.6 | 4.0 | .93* |
| 1214 | Reading-Literary | 3.7 | 4.0 | .92* |
| 1876 | Reading-Literary | 5.2 | 5.0 | .95* |
| | **Average for Category** | **4.4** | **4.4** | **.92** |
| 1614 | Reading-Informational | 3.4 | 3.5 | .76* |
| 1658 | Reading-Informational | 4.5 | 5.0 | .90* |
| 1676 | Reading-Informational | 5.6 | 5.5 | .95* |
| 1733 | Reading-Informational | 5.2 | 5.0 | .92* |
| 1963 | Reading-Informational | 5.4 | 5.5 | .94* |
| 1980 | Reading-Informational | 5.3 | 5.5 | .93* |
| 1992 | **Average for Category** | **4.9** | **5.0** | **.90** |

*Notes: Statistics based on 10 SMEs and rating scale where 1 = Not at all, 6 = Very well. *p < .05.*

Given that data from the rating approach can be aggregated and summarized for each of the dimensions comprising the test blueprints, we recommend this approach, which can be implemented by having SMEs review each item and rate the degree to which it appropriately measures the CCSS it is designed to measure. Based on the literature (e.g., O'Neil, Sireci, & Huff, 2004; Penfield & Miller, 2004), we recommend that at least 10 SMEs be used for each grade and subject area. This type of study will provide data that can be used to evaluate the content representativeness of items, sets of items that comprise an adaptive test for a student, and sets of items that comprise assessment targets, claims, or other levels of the test specifications. A contractor may propose a more general alignment study involving tasks that differ from those recommended here, which may be appropriate. However, the contractor should be required to demonstrate how the data will confirm the congruence between the sets of items that comprise an assessment for a student and the test specifications, as well as the degree to which the test items adequately represent the targeted cognitive skills. Although the adaptive nature of the summative assessments makes aggregating content validity results to a test "form" impossible, the representativeness of the most common sets of items taken by examinees, or a representative sample, could easily be studied (e.g., Crotts, Sireci, & Zenisky, 2012; Kaira & Sireci, 2010).

The content validity studies should also break out the results by item format. The summative assessments will include traditional selected-response items, technology-enhanced items, and performance tasks. Ideally, all item formats should have high ratings.

There is one drawback to the content validation/alignment methods discussed so far. By informing the SMEs of the CCSS measured by the items or of the assessment targets measured, they may exhibit a "confirmationist bias" or social desirability. That is, the SMEs may unconsciously rate the items more favorably than they actually perceive them, to please the researchers. One way around this problem is to have SMEs rate the *similarity* among pairs of test items and use multidimensional

scaling to analyze their data (D'Agostino, Karpinski, & Welsh, 2011; O'Neil et al., 2004; Sireci & Geisinger, 1992, 1995). However, this approach is not very common because it takes more time for SMEs to complete and involves more complex data analysis. A description of this method appears in Appendix C, should concerns about confirmationist bias/social desirability in evaluating test content arise.

*Evaluating evidence-centered design.* The evidence-centered design (ECD) underlying the development of the summative assessments specifies four claims and accompanying rationales in each subject area. These claims represent the cognitive models for each subject area. The assessment targets provide the evidence to support the claims, and the score reports represent the interpretation of the evidence. The content validity studies previously described could be extended to evaluate these three components of ECD in each subject area. The survey of curriculum specialists described earlier could include questions regarding the soundness of the claims and accompanying rationales in each subject area. Second, the studies involving ratings of items could be aggregated at the assessment target level to ensure that each target is represented by a sufficient number of items that are rated as measuring their intended CCSS well.

The third aspect of ECD, interpretation, should be evaluated through studies regarding the utility and comprehensibility of the summative assessment score reports. Ideas for these studies are described later in this report, in sections regarding validity evidence based on testing consequences. The idea here is to discover whether users of test reports interpret them correctly (Haertel, 1999), as well as if there are means for improving these score reports. It is assumed that studies of this kind will be done via piloting of the score reports. However, studies of the utility of the score reports should include ascertaining whether the information in the score reports is readily interpretable with respect to the intended claims.

**Validity Studies Based on Internal Structure.** Validity studies based on internal structure should be conducted to support the interpretations made on the basis of scores from the summative assessments. The scores reported should demonstrate adequate reliability and confirm the hypothesized "dimensionality" of the assessment. Studies in this area will involve analyzing the data from students' responses to the items.

*Dimensionality assessment.* With respect to dimensionality, it is presumed that items comprising the summative assessments will be calibrated using unidimensional IRT models, which are the most common models in contemporary educational assessment. One straightforward way to assess the dimensionality of tests calibrated using IRT is *residual analysis* (Hambleton, 1989; Hambleton & Rovenelli, 1986). Residual analysis compares the probability of success on an item (predicted by the IRT model) for students of different proficiency levels to the actual success of students of different proficiency levels.

Two examples of residual analysis plots are presented in Figures 5 and 6. The small circles in each figure are "conditional *p*-values" and represent the proportion of students, within a certain test score interval, who correctly answered the item. That is, they are proportion-correct statistics, conditional on test score (actually, conditioned on the IRT estimate of true score, called theta). The vertical lines spreading from these conditional *p*-values illustrate the confidence intervals for the probability estimates based on the IRT model. The item displayed in Figure 5 displays good fit, in that the IRT model for this item essentially runs through the conditional *p*-values. The item displayed in Figure 6 does not fit well, as several of the conditional *p*-values are far off the item characteristic curve specified by the IRT model.

Inspection of residual plots is descriptive in nature, and there are statistical indices that can be used to flag items that do not fit the IRT model. Such analyses are important for the summative assessments, to make sure that the various item types used are all adequately fit by the IRT model. More importantly, however, summary statistics across all items can be used to evaluate the degree to which the IRT model fits the data for all items comprising an assessment, and hence the degree to

which the IRT assumption of unidimensionality holds (note that a lack of fit may indicate a problem other than multidimensionality). All of the aforementioned analyses can be conducted using customized software, or the free ResidPlots2 residual analysis software developed by Liang, Han, and Hambleton (2008, 2009).[2] The ResidPlots2 software allows users to simulate data that fit the IRT model, to gauge the degree to which the observed test data deviate from chance expectations, assuming the IRT model is true. This analysis can be useful for evaluating overall IRT model fit to the data. Further description of ResidPlots 2 appears in Appendix D.

It should be noted that most IRT software programs produce residual plots and statistical measures of fit, such as the chi-square statistic. If the Smarter Balanced assessments were calibrated using the Rasch model, the Infit and Outfit measures of item fit could also be used to evaluate IRT model fit (e.g., Linacre, 2004).[3]

---

[2] Available for free from the University of Massachusetts at http://www.umass.edu/remp/software/residplots/.

[3] Both Infit and Outfit summarize the residuals between a student's observed pattern of responses to a set of items and the pattern predicted from the IRT model. The difference between the two measures is that the Infit measure weights items "closer" to a student's proficiency (theta) score more heavily than items further from the student's proficiency, whereas the Outfit statistic does not involve weighting. Each statistic represents a mean square error of the residuals and each has a standardized version.

Figure 5. IRT Residual Analysis Plot from ResidPlots-2 (good model fit)
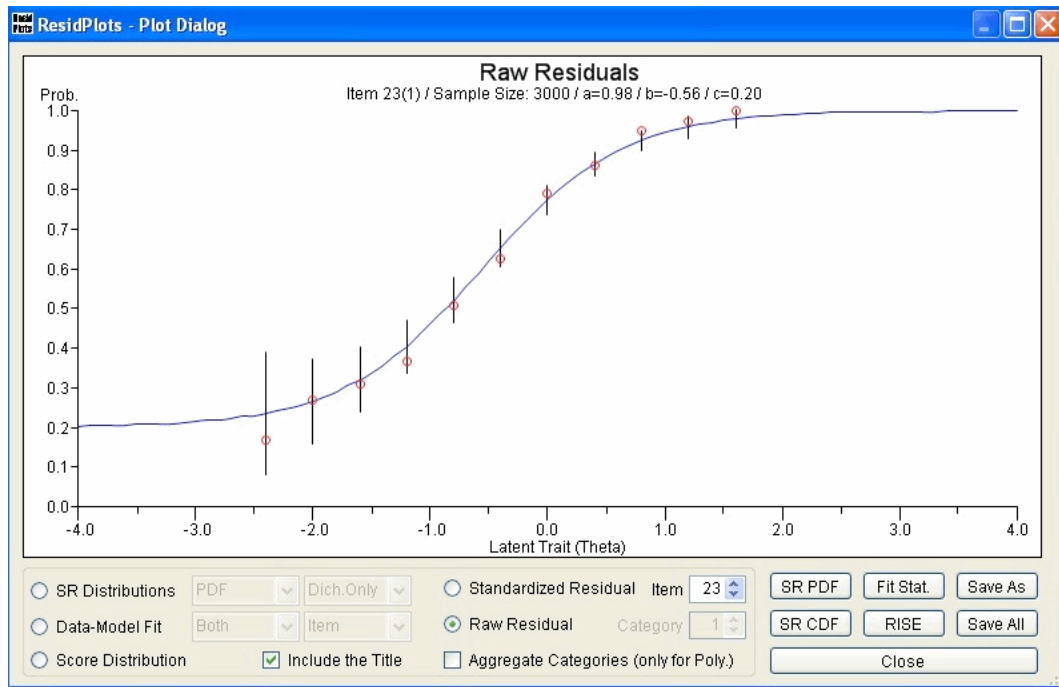


Figure 6. IRT Residual Analysis Plot from ResidPlots-2 (poor model fit)
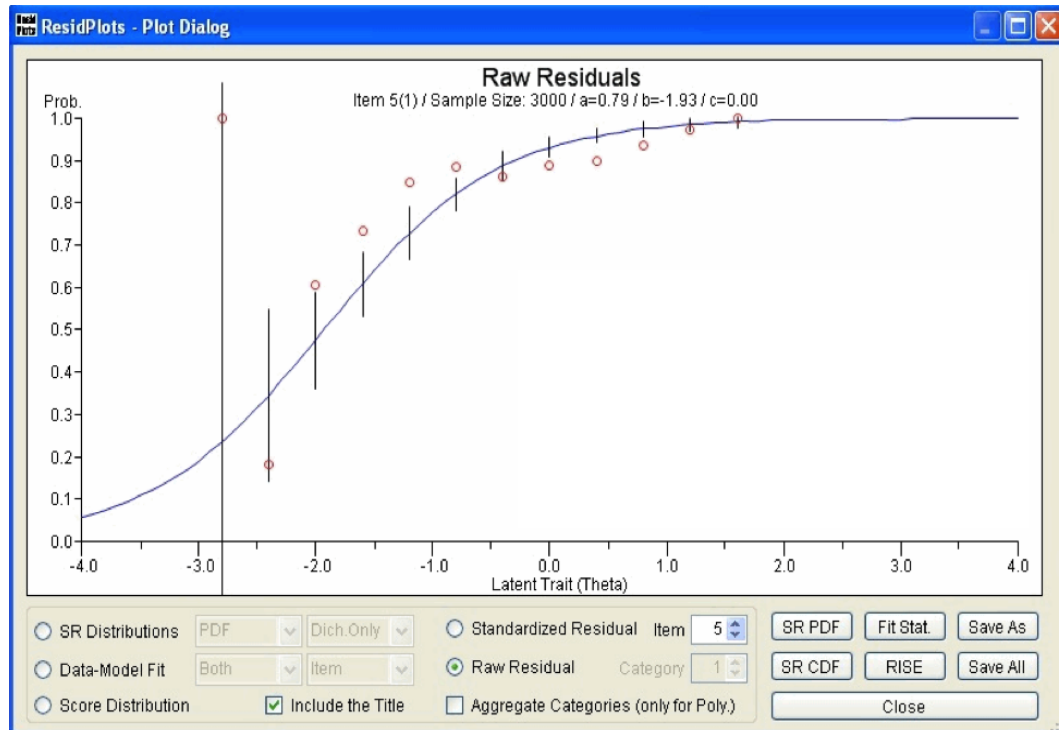
There are more comprehensive methods for assessing the dimensionality of an educational assessment, such as exploratory and confirmatory factor analysis and multidimensional scaling (see Hattie, 1985, or Sireci, 1997, for reviews of methods). Some of these methods are recommended for validity studies related to other Smarter Balanced purposes. For purpose 1, which is focused on whether the assessments are valid and reliable measures of the CCSS, evaluating dimensionality via residual analysis should be sufficient. An advantage of IRT residual analysis is that it can be easily conducted on "incomplete" data sets that result from adaptive testing—that is, the student-by-item data file is incomplete in that not all students respond to all items. Such nonrandom, missing data is difficult to analyze using standard factor analytic procedures (cf. Sireci, Rogers, Swaminathan, Meara, & Robin, 2000).

*Measurement precision.* Purpose 1 for the summative assessments specifies reliable measures, which involve an analysis of the precision of the assessments. *Measurement precision* refers to the amount of error, or variation, expected in a student's test score if the student were repeatedly tested. It is closely related to test score *reliability*, which is an estimate of the consistency or stability of the score. As described by Anastasi (1988):

> Reliability refers to the consistency of scores obtained by the same persons when reexamined with the same test on different occasions or with different sets of equivalent items, or under other variable examining conditions. This concept of reliability underlies the computation of the *error of measurement* of a single score, whereby we can predict the range of fluctuation likely to occur in a single individual's score as a result of irrelevant, chance factors. (p. 109)

*Measurement precision* is a broader term than *reliability* and refers to both estimates of score reliability and other descriptions of measurement error. A great deal of statistical theory has been developed to provide indices of the reliability of test scores as well as measures of measurement error throughout the test score scale. Classical test theory defines reliability as the squared correlation between observed test scores and their unbiased values ("true scores"). Reliability indices typically range from 0 to 1, with values of .80 or higher signifying test scores that are likely to be consistent from one test administration to the next.

Reliability indices are based on "classical" theories of testing. These estimates are reconceptualized in IRT, which characterizes measurement precision in terms of test information and conditional standard error. Therefore, the recommended measurement precision studies to support purpose 1 include estimates of score reliability (both coefficient alpha and stratified alpha, where relevant) and analysis of conditional standard errors of measurement based on IRT (e.g., test information functions and standard-error functions). Estimates of decision consistency, decision accuracy, and generalizability studies will be discussed in the sections related to other study purposes.

**Validity Studies Based on Response Processes.** The CCSS specify a wide range of knowledge and skills in each subject area. For example, two standards in high school geometry are:

> Know precise definitions of angle, circle, perpendicular line, parallel line, and line segment, based on the undefined notions of point, line, distance along a line, and distance around a circular arc.

and

> Construct an equilateral triangle, a square, and a regular hexagon inscribed in a circle. (NGA Center & CCSSO, 2010, p. 76)

The first standard represents a lower cognitive level of knowledge, while the second represents a higher level involving synthesis of several geometrical concepts. Evidence based on students' response processes could help validate that the summative assessment items are measuring the lower- and higher-order cognitive skills specified in the CCSS. One relatively easy study that could be done is an analysis of the amount of time it takes students to respond to items of various (purported) cognitive complexity. Students' response-time data should be readily available after the pilot tests, and the hypothesis that the items measuring higher-order skills will take more time for students to complete could be tested using analysis of variance (ANOVA).[4] In addition, cognitive interviews or think-aloud studies could be conducted to best understand students' thought processes as they respond to items of varying cognitive complexity (Hamilton, 1994; Leighton, 2004).

**Summative Assessment Purposes 2 and 3:**

> Provide valid, reliable, and fair information about whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA and mathematics to be on track for achieving college readiness.

> and

> Provide valid, reliable, and fair information about whether grade 11 students have sufficient academic proficiency in ELA and mathematics to be ready to take credit-bearing college courses.

These two purpose statements reflect the fact that the Smarter Balanced summative assessments will be used to classify students into achievement levels. Before grade 11, one achievement level will be used at each grade to signal whether students are "on track" to college readiness. At grade 11, the achievement levels will include a "college and career readiness" category. Such classification decisions require validation. Validity evidence for these purposes should come from four sources—test content, internal structure, relations with external variables, and testing consequences. In addition, because these classification decisions represent achievement level standards, Kane's (1994) sources of validity evidence for standard setting—procedural, internal, and external—are also relevant. However, we note that Kane's external evidence overlaps considerably with validity evidence based on relations with external variables.

Summative assessment purposes 2 and 3 differ with respect to grade level, with the assessments prior to grade 11 being used to predict whether students are "on track" for college and career readiness, and the grade 11 assessments used for certifying certain academic aspects of college and career readiness. This difference involves somewhat different types of validation evidence. In particular, because there has been a great deal of work on assessing college readiness, there are more potential validation criteria for the grade 11 college readiness classification.

---

[4] Note that response-time data are typically highly positively skewed, and so a natural log or similar transformation would be needed for this analysis.

**Validating "On Track" Based on Content Validity Evidence.** Being on track for college readiness implies acquisition of knowledge, and mastery of specific skills, thought to be important as students progress through elementary, middle, and high school. These specific knowledge and skills are stipulated in the CCSS. Therefore, the validity studies described earlier for purpose 1 are all relevant here. Essentially, the validity studies based on test content that were described for purpose 1 need to confirm that the summative assessments are targeting the correct CCSS and adequately represent these standards. However, such studies will not confirm that the CCSS actually contain the appropriate knowledge and skills to support college and career readiness. Rather, the CCSS would need to be reviewed to confirm that they contain the appropriate knowledge and skills that students need in order to be on track for college and career readiness.

One way to evaluate the appropriateness of the CCSS for determining whether students are on track for college and careers is to conduct a survey of state educators. At the postsecondary level, Conley, Drummond, Gonzalez, Rooseboom, and Stout (2011) conducted a national survey of postsecondary institutions to evaluate the degree to which the grade 11 and grade 12 CCSS contain the knowledge and skills associated with college readiness. They found that most (of almost 2,000) college professors rated these CCSS as highly important for readiness in their courses. A similar type of survey of educators in participating states would be helpful for evaluating the CCSS in ELA and math in grades 3 through 8. A major question motivating the survey would be: Are the CCSS in these grades appropriate for preparing students for college and careers?

In addition to these studies, it should be noted that studies involving validity evidence based on relations with other variables will also require validity evidence based on test content. For example, when Smarter Balanced assessment scores are compared with other test scores, the similarity of content across the two tests will need to be assessed.

**Validating "On Track" Based on Internal Structure Evidence.**

*Decision consistency and decision accuracy studies.* Given that purpose 2 involves the achievement level classification of "on track," in addition to the measurement precision studies described earlier for purpose 1 (IRT residual analysis, reliability estimates, information functions, etc.), evidence that the *classifications* assigned to students are reliable is needed. Therefore, estimates of decision consistency (DC) and decision accuracy (DA) are needed, as are estimates of the precision of measurement around the "on track" cut score (i.e., conditional error of measurement at that point).

In essence, DC refers to the consistency of student classifications resulting either from two administrations of the same examination or from parallel forms of an examination. Thus, the concept is similar to reliability, but instead of consistency of a score, it refers to consistency of classifications across repeated testing. DA can be thought of as the extent to which the observed classifications of students agree with the students' "true" classifications. Estimates of DA compare the classifications into which students are placed based on their test score with estimates of their true classifications. However, because students' true proficiencies are never known, simulation studies or some type of split-half estimate are typically used to estimate DA.

There are several statistical approaches for estimating DA and DC. Livingston and Lewis (1995) introduced a method for estimating DC and DA based on a single administration of a test, using classical test theory. More recently, IRT-based methods have been proposed (Lee, 2008; Rudner, 2001, 2004) and are more common for IRT-based tests. Free software for estimating DC and DA for IRT-based tests, such as the Smarter Balanced summative assessments, is available (Lee, 2008),[5] although some adjustments may need to be made for the adaptive test design. Another option would be the approach used by Hambleton and Han (2004), who estimated DA and DC by simulating data

---

[5] This software, IRT-Class, is available for free from the University of Iowa via http://www.education.uiowa.edu/centers/docs/casma-software/IRT-CLASS_v2_0_for_PC.zip?sfvrsn=0.

based on IRT item parameter estimates, and by comparing the consistency of classification over simulated examinees.

*Estimating the cut-score standard error.* As Kane (1994, 2001) discussed, analysis of the expected amount of variability in the cut score resulting from a standard setting study should be considered in validating an achievement level standard. As part of the documentation for setting the "on track" standard and other achievement level standards on the summative assessments, estimates of cut-score variability should be provided. These descriptive statistics estimate the amount of change expected in a cut score if the study were replicated using different panelists, items, or standard setting methods. Sireci et al. (2009) provided examples of several different methods for evaluating the cut scores established on a grade 12 National Assessment of Educational Progress (NAEP) mathematics assessment. These methods range from simply computing the standard error of the mean across panelists to replicating the standard setting study using an independent standard setting panel.

For the "on track" college readiness standards below grade 11, estimates of cut-score variability should be documented, but should also be communicated to Smarter Balanced leadership *before* the cut scores are finalized. The specific estimates to be used are somewhat dependent on the standard setting method. Most methods involve cut-score recommendations for each panelist, and so the standard error of the panelist mean can be computed. Where multiple rounds of standard setting are conducted in a study, the variability (e.g., standard deviation, standard error of the mean) across rounds can be calculated, with the expectation that variability will decrease across rounds.[6] When the panelists' median cut score is used, standard errors for the median can be computed based on bootstrapping (e.g., Sireci et al., 2009) and other procedures.

A better estimate of cut-score reliability is based on the variability across independent standard setting panels. Brennan (2002) showed that when there are only two independent observations, such as two means from two separate standard setting studies, the standard error of the mean is

$$\hat{\sigma} = \frac{|X_1 - X_2|}{2}$$

where $X_1$ and $X_2$ are the means across panelists in the two standard setting studies. For Smarter Balanced summative assessments that involve high-stakes standards, we recommend that independent standard setting studies be conducted so that the variability across recommended cut scores can be estimated.

**Validating "On Track" Based on Relations with External Variables.** It is likely that one of the achievement level standards set on the ELA and Math summative assessments will be used as the "on track" designation in each grade level. For example, the "Proficient" standard in each grade might be used. Validating this specific score interpretation based on the relations of scores with other variables requires other measures of students' mastery of grade-level knowledge and skills. Examples of external variables that could be used are teachers' ratings of students' preparedness for the next grade and other standardized assessments. Welch and Dunbar (2011), for example, explored the use of the Iowa Tests of Basic Skills (ITBS) for determining college readiness from grades 5 through 11. To accomplish this task, they first explored the relationship between the ITBS and the ACT composite scores for students who had taken the ITBS across grades and who had taken the ACT. The correlations between ITBS scores and the ACT ranged from .82 to .87 from grades 5 through 11. Next, for grade 11, they found the ITBS score that maximized classification

---

[6] Although computing statistics such as the standard error of the mean is common in standard setting studies, when panelists discuss their ratings, the independence-of-observations assumption is violated, and so this estimate of variability probably underestimates the true variability across independent panelists.

congruence with the ACT college readiness benchmark score (their study involved students who took both assessments). Using the corresponding ITBS percentile rank scores at the lower grade levels, they found about an 80% accuracy rate for predicting the ACT benchmark. However, they suggested putting error bands around the "on track" benchmark, and if a student's score was within the error band, the student could be considered on track.

In addition to the Welch and Dunbar (2011) study, both ACT and the College Board are using assessments at lower grade levels to assess college readiness. ACT has readiness benchmarks on its EXPLORE and PLAN assessments for grades 8 and 10, and the College Board recently introduced the ReadiStep exam for grade 8 and has long used the PSAT in Grade 10. The ACT benchmarks for EXPLORE and PLAN were set by retrospective analysis of students who took EXPLORE, PLAN, and the ACT.

Another study that could be conducted is to have teachers classify their students regarding whether each student is prepared for the knowledge and skills to be taught at the next grade level. Although subsequent-grade-level preparedness is different from college readiness, it is likely that these two variables would be strongly related. Thus, the classification consistency between teachers' ratings and students' "on track" classifications could provide useful validity evidence. For this type of study, teachers would have to be familiar with the curricula taught in the subsequent grade. We also recommend gathering data on teachers' *confidence* in the rating that they make for each student. Such data would be an important validity check before computing classification consistency and could be used to delete the data for teachers who were not confident in making their preparedness ratings for some or all students.

**Validating "On Track" Based on Testing Consequences.** Providing "on track" and other achievement level classifications for students in grades 3–8 is likely to have consequences for students, teachers, and instruction. At the student level, one potential negative consequence is promoting low academic self-esteem for students who are classified as below "on track." Such negative feelings could lead to "self-fulfilling prophecies" where students begin to believe that they are not smart or not capable of graduating high school. Student surveys and tracking dropout rates over time (Rabinowitz, Zimmerman, & Sherman, 2001) are two ways that this and other consequences could be measured. The "on track" designation could also have the intended positive consequence of early identification and remediation of students classified as below "on track." Therefore, following up on the instructional decisions that are made for these students is another area of study that would provide important validity evidence. Validity evidence for this purpose based on testing consequences should also involve gathering data from teachers via interviews, focus groups, or surveys to assess their perceived utility of these classifications and how it has affected their instruction. The consistency of these impressions and effects on instruction across grades should be studied.

**Validating "On Track" Based on Procedural Evidence.** Procedural evidence for standard setting refers to documentation and justification of all of the decisions and actions associated with a standard setting study. These decisions and actions were previously summarized in Table 5 (pp.20–21), and include selection of the standard setting panelists, justification of the standard setting method, training of panelists and other tasks associated with successful implementation of the method, analyzing the data, and assessing panelists' confidence in their ratings and the process. Justification of the standard setting method will be important for the Smarter Balanced assessments, as some methods, such as the widely used Bookmark method, have been shown to have serious deficiencies (Davis-Becker, Buckendahl, & Gerrow, 2011; Reckase, 2006a, 2006b). Procedural evidence must be comprehensively documented, and should include surveys of panelists and others involved in the process. Standard setting reports for NAEP, such as those by ACT (2005a, 2005b, 2005c) are excellent examples of comprehensive documentation of standard setting that provides procedural, internal, and external validity evidence.

### Validating College and Career Readiness Benchmarks

The third purpose statement for the summative assessments specifies college and career readiness. For the purposes of this research agenda, we assume that the knowledge and skills associated with college and career readiness have substantial overlap, as suggested by recent research (e.g., American Diploma Partnership, 2004; ACT, 2006), and so we focus on validating the college readiness benchmark. However, this assumption is based on convenience rather than research, since others have argued that the benchmarks for college and career readiness will be very different (Camara, in press; Loomis, 2011). Nevertheless, the methods described here for validating college readiness would carry over to the validation of career readiness, should appropriate external criteria for career readiness be identified.

**Validating "College and Career Ready" Based on Content Validity Evidence.** Up to this point, we have twice discussed validity evidence based on test content—first for purpose 1, and second with respect to students being "on track" for college readiness (purpose 2). The same studies apply here for validating the "college and career ready" inference based on the grade 11 summative assessments. This readiness designation implies acquisition of knowledge, and mastery of specific skills, considered necessary for success in college and careers and stipulated in the CCSS. Therefore, the content validity studies described earlier for purpose 1 are relevant here, and their findings should inform the validity argument for validating the college and career readiness standard. The additional evidence required for readiness is evidence that these standards are, in fact, the appropriate prerequisite skills in math and ELA that are needed to bypass remedial college courses and be ready to successfully begin postsecondary education or a career. The recent report by Conley et al. (2011) represents important evidence to support that assumption. Similarly, Vasavada, Carman, Hart, & Luisser (2010) found strong alignment between College Board assessments of college readiness and the CCSS.

Other validity evidence that is based on test content and that will be used in the validity argument for the college and career readiness determination includes content overlap (alignment) studies that will be done to gauge the similarity of knowledge and skills measured across the summative assessments and external assessments that are used to evaluate the readiness standards. Postsecondary admissions tests (e.g., ACT, SAT) and college placement tests (e.g., ACCUPLACER, AP, Compass) will be used in concurrent and predictive validity studies, and so the overlap of skills measured must be documented to properly interpret the results. The National Assessment Governing Board (NAGB) recently began a program of research in this area to set college and career benchmarks on the grade 12 NAEP assessments. Its research agenda began with comprehensive alignment studies that evaluated the overlap of NAEP and external assessments (Loomis, 2011; NAGB, 2010).

**Validating "College and Career Ready" Based on Internal Structure Evidence.** The previous descriptions of validity evidence based on internal structure for the "on track" student classification (i.e., estimates of DC and DA, review of the conditional standard error of measurement around the cut score, estimates of the standard error of the cut scores derived from the standard setting studies) are equally important for validating the college and career readiness classifications of students. These estimates and studies were described in previous sections, and so their descriptions are not repeated here.

**Validating "College and Career Ready" Based on Relations with Other Variables.** In considering validating the college readiness achievement level standards on the Smarter Balanced summative assessments, we focus on validity evidence based on relations to external variables because, as Camara (in press) pointed out, "Given the intended purposes of [college and career readiness] assessments, if performance levels and benchmarks are inconsistent with empirical data of performance in college and career-training programs, they will not only lack credibility but would raise concerns about the validity of the interpretive argument."

A college- and career-ready standard implies that students who meet this standard have the prerequisite academic knowledge and skills to succeed in college or in a career. Given that there are currently existing standards for college readiness,[7] the readiness classifications based on the Smarter Balanced summative assessments should be congruent with these other standards, assuming that these external standards accurately measure college readiness. The degree to which current college readiness benchmarks are consistent with the Smarter Balanced readiness standards needs to be studied. These studies could be used (a) to empirically set the Smarter Balanced readiness standards, (b) as part of the standard setting process, or (c) to validate the standards after they have been set by other means.

Validity evidence based on relations to other variables for the purpose of classifying students as college ready should involve both correlation/regression studies and classification consistency analyses. In these analyses, scores from the summative assessments will be correlated with, used as predictors of, and cross-tabulated with other measures of college readiness. To conduct these analyses, appropriate external measures must be identified, defined, and evaluated for validation purposes. In addition, different research designs should be considered. Design options include:

- Concurrent studies where students take both the summative assessments and external assessments;

- Predictive studies where students take the summative assessments and their future college performance is compared in retrospective fashion; and

- Embedded item designs where summative assessment items are embedded in other assessments of college success, and vice versa.

Defining "college success" is not straightforward, and so we recommend that several different variables be used, and studied, as outcome variables for college readiness. Camara (in press) listed seven criteria that have been or could be used for setting or evaluating college readiness benchmarks on Smarter Balanced or Partnership for Assessment of Readiness for College and Careers (PARCC) assessments. These are:

- Persistence to second year;

- Graduation or completion of a degree or certification program;

- Time to degree completion (e.g., 6 years to earn a bachelor's degree);

- Placement into college credit courses;

- Exemption from remediation courses;

- College grades in specific courses; and

- College grade point average.

Camara also noted that the most common criterion is college grades, either first-year grade point average (GPA) or grades in specific first-year courses. For example, in setting the college readiness benchmark on the ACT, grades in specific first-year courses were used (Allen & Sconing, 2005), but to set the same benchmark on the SAT, Wyatt, Kobrin, Wiley, Camara, and Proestler (2011) used first-year GPA.

---

[7] We use *readiness* here to refer to the academic skills in math and reading, not the more general readiness criteria that include non-cognitive variables such as contextual skills and academic behaviors (Conley, 2007).

*Current college readiness benchmarks set on educational tests.* Several studies have been used to evaluate or set college readiness benchmarks on tests. Examples of testing programs that have set or evaluated college readiness benchmarks include:

- ACCUPLACER
- ACT
- Advanced Placement exams
- COMPASS
- Current statewide high school tests (end-of-course or graduation tests)
- Early Assessment Program (California)
- EXPLORE
- International assessments (e.g., PISA, TIMSS)
- International Baccalaureate
- NAEP
- PLAN
- PSAT/NMSQT
- ReadiStep

A recent report by NAGB (Fields & Parsad, 2012) found that the most common assessments used by postsecondary institutions to evaluate entering students for remedial courses in math were the ACT, SAT, ACCUPLACER (Elementary Algebra and College Level Math), and COMPASS (Algebra, College Algebra). For reading, the most common assessments were the ACT, SAT, ACCUPLACER (Reading Comprehension), ASSET (Reading Skills), and COMPASS (Reading).

Examples of some of the studies that have been done using these tests, the readiness standards that were set on each, and relevant citations are presented in Table 6. Camara (2012) described research in this area as consisting of three steps: First, determine the appropriate outcome variable for college success (e.g., first-year GPA). Second, determine the appropriate criterion of "success" on the outcome variable (e.g., 65% chance of a B-). Third, determine the appropriate probability of success. These steps will be important considerations in designing validity studies for the Smarter Balanced summative assessments.

Table 6. Current College Readiness Benchmarks

| Test | Criterion | Benchmark | Comments/Citations |
|---|---|---|---|
| ACT English | .75 probability of C and .50 probability of B | 18 | Allen & Sconing (2005) |
| ACT Reading | | 21 | |
| ACT Math | | 22 | |
| SAT Composite | .65 probability of B- in first-year GPA | 1550 | Wyatt et al. (2011) |
| SAT-Quantitative | | 500 | |
| SAT-Reading | | 500 | |
| SAT-Writing | | 500 | |
| Advanced Placement (AP) | | Score of 3 | Relevant tests include Calculus AB, Calculus BC, English Language & Composition, English Literature & Composition, and Statistics. |
| COMPASS | .75 probability of C and .50 probability of B | 77 (English), 52 (Math) | ACT (2010) |
| EXPLORE | .75 probability of C and .50 probability of B | 13 (English), 17 (Math) | ACT (2010) |
| PLAN | | 15 (English), 19 (Math) | ACT (2010) |

The studies reported in Table 6 primarily used regression methods to find the test score that best distinguished students who met or did not meet some operationally defined criterion of college success.[8] For the ACT research, the criterion used was the test score associated with a .75 probability of earning a C or a .50 probability of earning a B in specific college courses (e.g., English composition, college algebra). For the SAT research, the criterion used was the test score associated with a .65 probability of earning an overall first-year GPA of B- (2.67). The ACT studies used linear regression, whereas the SAT studies used logistic regression. The SAT studies also included validity evidence based on external variables, specifically rigor of high school courses, AP exam scores, and high school GPA, to support the SAT readiness benchmarks (Wyatt et al., 2011). In addition to the studies reported in Table 6, Fields and Parsad (2012) conducted a comprehensive survey of cutoff scores on postsecondary math and reading placement tests. The mean cutoff scores, and the variability in these scores across institutions, were reported. These mean cutoff scores could be used as validation criteria for the Smarter Balanced college readiness standards. Other readiness criteria include specific cutoff scores used by state university systems (e.g., California and Texas have readiness criteria based on the ACT, the SAT, and in-state assessments), and the International Baccalaureate exams (compensatory score of 24 across six assessments).

---

[8] Equipercentile equating could also be used, and may be preferable in some situations.

In addition to establishing college readiness benchmarks on admissions tests, research has also been conducted to see how these readiness benchmarks could inform setting readiness standards on other assessments. For example, the Texas Education Agency commissioned a series of studies to set and evaluate college readiness standards using the State of Texas Assessments of Academic Readiness (STAAR). In fact, in establishing the new STAAR tests, the Texas legislature legislated that "validity studies be conducted to evaluate the empirical links between student performance on the STAAR assessments and specific assessments measuring similar constructs, and that these links be used to inform the standard-setting process" (LaSalle et al., 2012, p. 2). These studies are particularly relevant to Smarter Balanced because the STAAR assessments involve on-target readiness standards below high school and certifying college readiness at the high school level.

Rather than directly using external assessments to set readiness benchmarks on the STAAR exams, Texas used external data to set "landmarks," or cut points, on the STAAR score scale that corresponded to important cut scores on the external assessments. Examples of external assessments that were used for this purpose included the previous statewide exams in Texas, a placement test used at the University of Texas, the ACT and SAT benchmarks, and the ACCUPLACER Elementary Algebra exam. For the previous statewide end-of-course tests, equipercentile linking was used to establish concordance tables across pairs of tests. For the readiness benchmarks established on the external assessments, logistic or linear regression was used to "map" the external benchmarks onto the STAAR score scales. Linear regression was also used to set other landmarks based on high school course grades (e.g., B or better) and probability of success in a relevant college course (e.g., C or better in college algebra). See Keng, Murphy, and Gaertner (2012) for a more complete description of these studies.

Based on several studies of these external criteria, "landmarks," or benchmarks, were established on the STAAR score scale, and these landmarks were used to establish "neighborhoods" within which it seemed reasonable (to the policymakers who reviewed these results) to set the college readiness standard and other standards. The score scale annotated with the landmarks and neighborhoods was used to encourage standard setting panelists to set their standards within the neighborhoods, since the score scale range defined by each neighborhood contained the external readiness standards and other relevant information that would support the standard set in that range. Keng et al. (2012) described this process as "evidence-based standard setting" (p. 4; see also O'Malley, Keng, & Miles, 2012).

A fictitious example of how external data could be used to inform the college and career readiness standard setting process using neighborhoods based on external data is presented in Figure 7. In this figure, test scores related to college readiness from two states (California and Oregon), the ACT and SAT readiness benchmarks, and the passing score for the GED Math test are all mapped onto the score scale for the grade 11 Smarter Balanced summative math assessment. The score corresponding to chance performance is also indicated. Using external data in this way can build validation criteria into the standard setting process.

Figure 7. Example of Using External Data to Establish a Reasonable Interval (Neighborhood) for Standard Setting
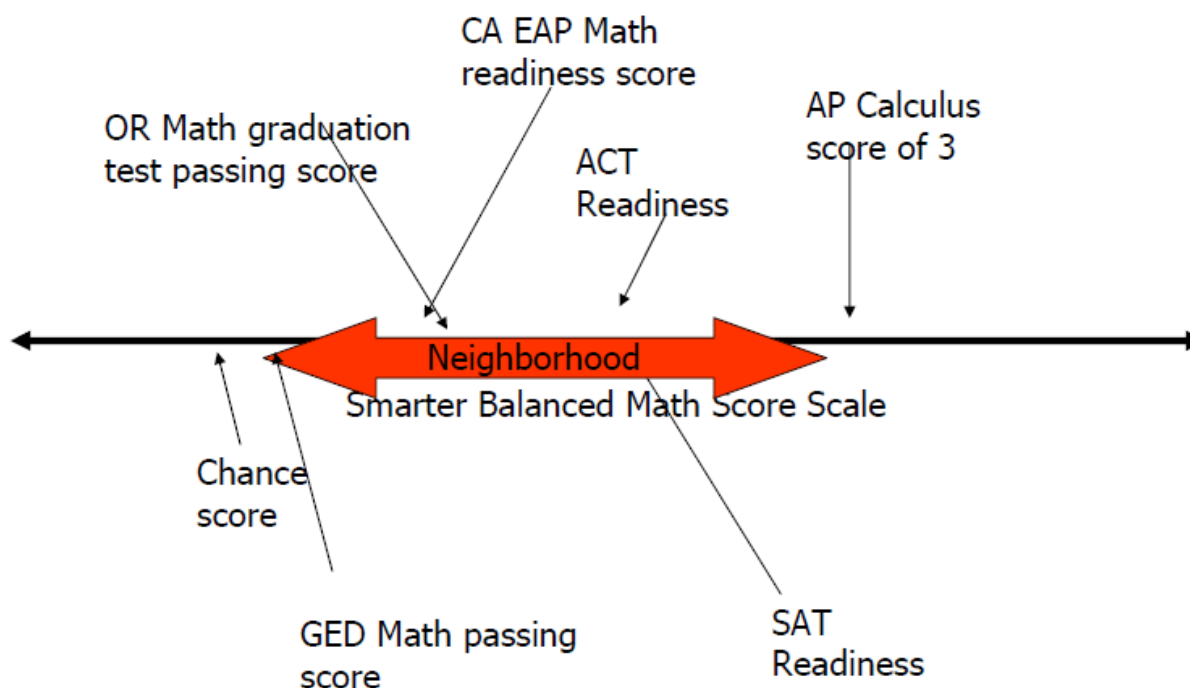


*Recommended studies based on relations to external variables.* The previous section described some options for conducting validity studies based on relations to external variables and summarized some of the research that has already been done in this area. To relate current college readiness standards and other pertinent information to the grade 11 Smarter Balanced summative assessments, three types of studies are possible. The first two types of studies are concurrent validity studies. In the first variation, students would take both Smarter Balanced and external assessments at around the same point in time. For example, grade 11 students could take the Smarter Balanced summative assessments, or a subset of items from them (e.g., in the pilot study), and the SAT or ACT, at a reasonable point in time (e.g., March). Regression or equipercentile methods could be used to determine the Smarter Balanced scores that corresponded to the SAT or ACT readiness benchmarks. The second type of concurrent validity study would involve college students taking Smarter Balanced assessments (or subsets of items) near the end of a relevant course, and their final course grades could be used as the validation criterion. The Smarter Balanced scores that are associated with the pre-established readiness criterion (e.g., grade of B-) could be established via regression or equipercentile procedures, or probability tables could be set up to relate the Smarter Balanced scores to specific grades. The third type of study that could be conducted would be a retrospective study where students who took the Smarter Balanced assessments would be followed longitudinally to see how they perform in college (see, for example, D'Agostino & Bonner, 2009).

Threats to the validity of these studies include differential motivation effects across the Smarter Balanced and external assessments, potentially non-representative samples of students due to the self-selection of external assessments, and a lack of overlap in the constructs measured by the Smarter Balanced and external assessments. Different grading standards and different admissions standards across colleges and universities, and across different types of institutions (public, private,

two-year, four-year) also present problems. Nevertheless, these issues can be considered and discussed when interpreting the results. Surveys or interviews of students participating in these studies could help understand these students' motivation to do well (Haertel, 1999).

The most practical course of action to gather external data to validate the Smarter Balanced college readiness standards is to take advantage of tests already taken by grade 11 students, such as the ACT, SAT, and AP exams, and relate them to their scores on the summative assessments. Supplementary studies would need to evaluate the content overlap of these assessments and students' motivation to do well on the Smarter Balanced assessments. Assuming sufficient content overlap and motivation, benchmarks can be set to inform the establishment of the college readiness standards on the Smarter Balanced assessments (as done in Keng et al., 2012), and longitudinal analysis can be done at a later point in time to evaluate the standards and possibly revise them if necessary. The key information to gather is the degree to which students who reached the Smarter Balanced readiness standards were successful in college. Camara and Quenemoen (2012) suggested that the decision consistency of the ready/not-ready and successful/not successful in college classifications should be broken down across different types of institutions.

It is likely that data-sharing agreements that maintain student anonymity can be worked out between the Consortium and external examination programs, such as ACT and the College Board, and among state colleges and universities within the Consortium. In addition, as Camara and Quenemoen (2012) point out, the National Student Clearinghouse maintains enrollment records for a vast majority of postsecondary institutions and can be used to track retention and graduation rates that will be useful for evaluating the readiness standards. The percentages of students who are "Proficient" on the grade 12 NAEP Math and Reading assessments will also be evaluated with respect to the percentages of students who are classified as "college ready" on the respective Smarter Balanced assessments. Should the NAEP grade 12 results ever be reported at the state level, within-state NAEP/Smarter Balanced comparisons would be informative.

**Validating "College and Career Ready" Based on Testing Consequences.** The college and career readiness standard on the Smarter Balanced summative assessments is intentionally integrated with the "on track" standards set at the lower grade levels. The intended consequence of this system is better preparation of students so that they are prepared for college or careers by the time they graduate high school. This intended consequence can be measured by analyzing trends in college completion and remedial course enrollments over time, and by surveying secondary and postsecondary educators about students' proficiencies. However, validity evidence for the college and career readiness designation should also investigate unintended consequences, such as unanticipated changes in instruction, diminished morale among teachers and students, and increased pressure on students that may lead to dropout, or to pursuing college majors and careers that are less challenging. To evaluate these potential consequences, teacher surveys of enacted curriculum, student surveys of career aspirations, and psychological assessments of anxiety and academic self-concept could be conducted.

The recommended studies based on testing consequences that will target the college and career readiness purposes should include teacher surveys regarding changes in student achievement and preparedness over time and changes in teachers' instruction over time. We also recommend that students be surveyed regarding college and career aspirations. Student and teacher samples that are representative at the state level would suffice for these studies. If time and resources permit, assessing the anxiety levels of students regarding their likelihood of obtaining college or career readiness, and their academic self-concept, would also be helpful. Validity evidence based on the consequences of the college and career readiness standard should also involve analysis of secondary and postsecondary enrollment and persistence, changes in course-taking patterns over time, and teacher retention for teachers in math and ELA.

**Summative Assessment Purpose 4:**

> Provide valid, reliable, and fair information about students' annual progress toward college and career readiness in ELA and mathematics.

As indicated in Table 1, validity evidence to support the use of the summative assessments for providing information about students' annual progress should be based on test content, internal structure, relations with external variables, and testing consequences. Studies related to test content need to evaluate the degree to which similar standards are measured across grades and the consistency of the construct across grades. Studies based on internal structure should evaluate the validity of the vertical scale used to measure progress over time. Studies involving relations with external variables are needed to confirm that the progress observed on the Smarter Balanced scale is mirrored by other measures of academic achievement. Finally, studies based on testing consequences should confirm that the measures of annual progress have a positive effect on instruction and student learning.

The most straightforward way to measure changes in students' proficiencies over time is to have scores from assessments at different points in time on a common scale. The physical analogy is the bathroom scale that remains unchanged across different measurements of weights. Sometimes, however, even the bathroom scale needs to be recalibrated to confirm the zero point. With educational assessments, it is difficult to put scores from assessments at different time periods on the same scale, because the items administered to students at different points in time are not the same. At this juncture, the Smarter Balanced summative assessments are planned to be vertically equated across grades, which means that a single score scale will span the grades. A vertical scale facilitates measuring changes in students' performance over time (Briggs, 2012; Kolen, 2011; Patz, 2007). However, it is difficult to create a valid vertical scale. Challenges to vertical scaling include changes in the construct of math or ELA across grades, and differences in when material is taught across grades and schools (Tong & Kolen, 2007). Therefore, validity evidence to support measuring students' progress toward college and career readiness should involve evaluation of the vertical scale across grades.

**Validity Studies Based on Test Content.** Evaluations of the content measured across grades will be an important source of evidence for validating the appropriateness of the vertical scale for measuring students' progress. First, this evaluation should assess whether there is overlap among the CCSS measured across adjacent grades (Patz, 2007). Next, the evaluation should review the common items that are used to form the vertical links across grades. SMEs should be asked whether the linking items are relevant to students in both grades and if they adequately represent the expected learning progressions. The content review should also assess the degree to which a common construct can be considered to hold across grades, or at least across adjacent grades. For example, do the anchor items that are used across grades measure CCSS that are appropriate for each grade?

**Validity Studies Based on Internal Structure.** Most of the studies that should be conducted to evaluate the validity of the vertical scales underlying the summative assessments can be categorized as evidence of internal structure. These studies include dimensionality analyses and evaluation of item statistics, mean scores, and score distributions across grades.

*Dimensionality analyses.* One important area of study is evaluation of the dimensionality of the assessment data, and of the degree to which the dimensionality is consistent across grades, or at least across adjacent grades. For example, if a single dimension is hypothesized to exist across grades, the degree to which the data for each grade are unidimensional, and the degree to which the same dimension holds across grades, should be studied. One way to conduct this analysis is using IRT residual analysis, as suggested earlier. The added layer of analysis would be evaluating the consistency of the fit across grades. Kolen (2011) noted that "even if the unidimensionality assumption does not strictly hold, the IRT model might provide an adequate enough summary of the

data that the vertical scale is still useful" (p. 12). Other dimensionality assessment procedures, such as confirmatory factor analysis or bifactor analysis, could also be useful.

The incomplete student-by-item data matrix that results from adaptive testing can cause problems for many dimensionality assessment procedures, such as exploratory and confirmatory factor analysis. Thus, assessing the dimensionality within and across grades within an IRT framework is probably most practical. In addition to residual analysis, both unidimensional and multidimensional IRT models can be fit to the data, and the difference between models can be tested for significant and practical improvement in fit to the data (Bock, Gibbons, & Muraki, 1988; Sireci, 1997). For items that are dichotomously scored, this analysis can be conducted using the TESTFACT software (Wilson, Wood, & Gibbons, 1991). To assess multidimensionality using both dichotomous and polytomous items, some specialized software may be needed.

*Analysis of statistics across grades.* The establishment of a vertical scale implies an increase in the difficulty of the assessments as grade increases and higher proficiency of students in higher grades relative to lower grades. At the item level, it is assumed that students at a higher grade level will have a higher probability of correctly answering an item than students at a lower grade level. These assumptions can be checked to evaluate the validity of the scale. Factors such as when students are taught specific knowledge and skills (i.e., opportunity to learn) and difference in time between instruction and assessment can cause "reversals" where students at higher grade levels perform worse than students at lower grade levels. Such reversals can be a problem when a common item approach is used to link the assessments across grade levels. Therefore, an additional study is a comparison of where the items "land" on the vertical scale versus the grade levels for which they were written. For example, if items written for a grade 6 assessment have IRT difficulty estimates that put them in the general range of grade 5 or grade 7 items, there will be a disconnect between the intended content at each grade level and the actual scale properties.

Kolen (2011) and Patz (2007) suggested several analyses that could be used to evaluate the validity of a vertical scale (see also Kolen & Brennan, 2004). These analyses include:

- correlation of item difficulties across grade levels

- a progression in test difficulty of test characteristic curves across grades

- analysis of item difficulties across grades

- comparison of mean scores across grades

- comparison of scale scores associated with proficiency levels across grades

- comparison of overlap of proficiency distributions across grades

- comparison of variability in test scores within and across grades

Validity evidence for vertical scales that are appropriate for measuring students' annual progress would include a lack of reversals of item difficulties across grades, anticipated separation of means and proficiency distributions across grades, and sensible patterns of variability within and across grades. With respect to comparison of score means across grades, Patz (2007) suggested, "For sufficiently large and diverse samples of students, scale score means would be expected to increase with grade level, and the pattern of increase would be expected to be somewhat regular and not erratic" (pp. 17–18).

With respect to evaluating patterns of variability, Kolen (2011) noted:

> Within grade variability indices typically are either similar across grades or increase as grade increases. Either of these patterns seems reasonable. Sometimes within grade variability indices decrease substantially as grade increases, which is sometimes referred to as *scale shrinkage*. Scale shrinkage can be indicative of

problems with IRT parameter estimation, in which case the vertical scaling procedures might need to be adjusted or the scale abandoned. (p. 12)

In considering establishing a vertical scale for PARCC, Kolen noted, "PARCC might decide, based on the construct being assessed, that an acceptable vertical scale should display increasing mean scores from year to year, that the amount of growth is decelerating, and that the within grade variability is either approximately equal across grades or is increasing from grade to grade" (p. 21). These evaluation criteria are applicable to evaluation of the Smarter Balanced vertical scale for the summative assessments.

In addition to analyses of item statistics and test scores across grades, Briggs (in press) claims that vertical scales should be validated by demonstrating that they possess interval scale level properties. This idea is new and has not seen wide application, but Briggs suggests the use of additive conjoint measurement to determine whether vertical scales have equal-interval properties, which he considers necessary for valid measurement of students' annual progress

In addition to the previously mentioned studies, analyses of *item parameter drift* over time should also be conducted. These analyses involve recalibrating IRT item parameters in subsequent years and comparing them to their estimates in prior years. Such analyses could improve the anchors used in equating across years by eliminating anomalous items, or could identify items that have been compromised (i.e., security problems).

**Validity Studies Based on Relations to Other Variables.** To confirm that the summative assessments provide valid information about students' annual progress in math and ELA, it would be good to compare students' progress on these assessments with other measures of their achievement over the same time period. At a macro level, the aggregated progress of students over time could be compared to changes of students within a state on the NAEP math and reading assessments. On an individual student level, progress on the Smarter Balanced assessments could be compared to other standardized assessments that are on a vertical scale, such as the ITBS or the Measures of Academic Proficiency (Northwest Evaluation Association, 2005).

In addition to concurrent validity evidence based on other tests, the degree to which the summative assessments are sensitive to instruction could also be studied to evaluate the degree to which the tests measure students' annual progress. Teachers who more fully implement the CCSS into their instruction should have students who make greater progress on the summative assessments. D'Agostino, Welsh, and Corson (2007), for example, measured the degree to which teachers emphasized state academic standards in their teaching and compared these measures to students' performance on the statewide test. They found a modest but positive relationship. A similar strategy could be implemented to evaluate the patterns of progress noted across classes on the summative assessments. Another way in which external data can inform the validation of the summative assessments as a progress measure is to have teachers rate the math and ELA progress made by their students within a year, and compare it to their progress as measured by the Smarter Balanced score scales.

**Validity Studies Based on Testing Consequences.** The summative assessments are supposed to provide information regarding students' annual progress so that their progress toward college and career readiness can be ascertained. If adequate progress is not found, it is likely that instructional changes will be made to support improved progress. Thus, validity evidence based on testing consequences should include surveys or interviews of teachers to understand the degree to which they find estimates of students' progress helpful for targeting instruction to individual students and to their classes in general. In addition, if progress measures are used to alter the instruction for a student—for example, placing the student in supplementary instruction or an after-school program—the degree to which these actions are associated with improved progress should be studied (Shepard, 1993).

Another important study of testing consequences related to measuring progress is the degree to which progress is similar across subgroups of students. If students from different ethnic backgrounds, socioeconomic statuses (SES), or disability statuses are progressing at different rates, the reasons for such differential progress should be studied. It may be that students who initially perform low on the assessments have more opportunity to exhibit progress. In any case, patterns of progress across subgroups should be studied to ascertain whether these patterns are expected given the student characteristics, or if they reflect some insensitivity of the assessments to properly capture progress or some type of deficiency in the scale properties.

### Summative Assessment Purpose 5:

> Provide valid, reliable, and fair information about how instruction can be improved at the classroom, school, district, and state levels.

As indicated in Table 1, for the Summative Assessments to provide information that will improve instruction, the content of the assessment must adequately measure the intended CCSS, and teachers, administrators, and other educators must appropriately act upon this information to tailor instruction accordingly. The validity studies based on test content that were described earlier for purposes 1 through 4, and the studies of testing consequences that were described for purposes 2 through 4, would all provide evidence regarding the degree to which the assessment results are instructionally relevant. The gathering of additional validity evidence to support purpose 5 will be similar to the studies suggested later in this report for the interim assessments and formative assessment resources, because these components are designed to work together to improve instruction. Many of these studies fall under the category of validity evidence based on testing consequences; one study based on relations to other variables, which was already mentioned with respect to purpose 4 (a study of sensitivity of the summative assessments to instruction), is also relevant to purpose 5.

As noted earlier, teachers who more fully implement the CCSS into their instruction should have students who make greater progress on the summative assessments (D'Agostino et al., 2007).

**Validity Studies Based on Testing Consequences.** The provision of summative assessment information to improve instruction will most likely come from the score reports associated with these assessments. Therefore, the evaluation of testing consequences relative to this purpose will focus largely on the utility of these score reports. An analysis of classroom artifacts will also provide important evidence, as will the types of surveys, interviews, and focus groups associated with the studies mentioned earlier for purposes 1 through 4.

*Studies on effectiveness of summative assessment score reports.* According to the score reporting RFP (RFP-15), Smarter Balanced has planned a wide and comprehensive variety of score reports to support purpose 5. There will be both static score reports and dynamic score reports that are interactive. Summative assessment results will be reported at the total score and claim levels for both subject areas, and reports will be available for both individual students and aggregate groups. The comprehensive nature of these reports, and their online access and variety, should provide actionable data to improve instruction at the classroom, school, district, and state levels. Research studies should be conducted to confirm that these intended consequences are occurring.

RFP-15 requires gathering feedback from potential users as score reports are being developed. Documentation regarding these reports should be reviewed to see what changes were made on the basis of this feedback. In addition, once the reports are operational, studies should be conducted to ascertain how well teachers, administrators, parents, students, and other stakeholders (e.g., legislators, journalists) understand the reports and find them useful. These studies should include surveys, focus groups, and interviews. In addition to gathering stakeholders' impressions of the reports, their understanding of the information contained in the reports should be tested (Wainer, Hambleton, & Meara, 1999). The actions that teachers take based on the score reports should also

be documented and evaluated for appropriateness (Bennett, 2010). In addition to assessing users' understanding and use of the reports, surveys should also be used to inquire about ease of navigating the system, timeliness of data, and additional features that users would like to see.

Analyses of usage statistics should also be conducted to determine the most popular reports and to confirm that all reports created are being used. The different types of reports that users create should also be reviewed. The most commonly used and least commonly used reports could be targeted for discussion in focus groups to (a) ensure that users are making appropriate inferences from the reports, (b) ensure that taking appropriate actions based on the reports, and (c) discover how the least-accessed reports could be improved to make them more useful, or to make users aware of them.

To maximize utility of the reports, users or "data coaches" should be trained on how to access them and use them. In fact, the Peer Review Guidance (U.S. Department of Education, 2009b) stated that "Training on interpretation of results is required [and] must provide evidence on how educators can interpret results and then use them for proper decision making" (p. 69). Thus, the effectiveness of the training should also be evaluated.

*Studies of textbooks and classroom artifacts.* Another way in which the effects of the summative assessments on instruction can be evaluated is by looking at changes in textbooks and instructional practices before, during, and after implementation of the assessments. In addition to the surveys and interviews previously discussed, classroom artifacts such as lesson plans, student handouts, classroom assessments, homework, syllabi, and teacher logs (e.g., Silk, Silver, Amerian, Nishimura, & Boscardin, 2009; Tomlinson & Fortenberry, 2008) should be studied.

## Summative Assessment Purpose 6:

Provide valid, reliable, and fair information about students' ELA and mathematics proficiencies for federal accountability purposes and potentially for state and local accountability systems.

Results from the summative assessments will include scale scores at the total score and claim levels, and achievement level classifications in each subject area. The achievement level results could be used as they are currently employed in statewide testing programs for federal accountability purposes under NCLB. In addition, students' progress over time could be used in growth models for other accountability purposes, some of which may be for federal accountability and some at the state or local levels. The Smarter Balanced principle of "responsible flexibility" (Smarter Balanced, 2010, p. 5) is consistent with the idea of providing valid, reliable, and fair information that can be used for federal accountability in uniform fashion across all participating states, but also allows for states to use information from the summative assessments in their statewide and local accountability systems.

Smarter Balanced cannot assume the responsibility for validating all of the potential uses of the summative assessments at the state and local levels, but the responsibility for validating accountability at the federal level should be included in the research agenda. In particular, the metric of "percent proficient" at the total student population level and at the subgroup level should be validated, as well as any other aggregate statistics used for federal accountability.

Percent proficient is currently a primary accountability criterion in NCLB, which also requires states to set at least three proficiency levels. In considering the reporting of achievement level results in California, a technical advisory committee led by Lee Cronbach (Select Committee, 1994/1995) recommended that (a) the percent *above* cut points be reported, rather than percents at proficiency levels; (b) only one percent above cut points, or two at most, rather than percent above cut points for all proficiency levels, be reported; and (c) standard errors for percent above cut points be reported (Yen, 1997). The first two recommendations were suggested to reduce confusion in reporting scores

to the public. The third recommendation is standard practice in reporting scores for accountability or other purposes.

The provision of valid, reliable, and fair information has been covered in the previous purpose statements, through the various studies involving test content, internal structure, relations to other variables, response processes, and testing consequences. The additional studies needed to validate the accountability uses of Smarter Balanced summative assessment scores are studies involving the reliability and validity of *aggregate* scores used for accountability. Of particular importance is the reliability of aggregate scores.

**Studies Evaluating the Reliability/Precision of Aggregate Scores.** Individual schools will be one aggregate level of analysis in federal accountability, and so the reliability or error associated with school-level results will need to be estimated as part of the validity research agenda. If accountability results will be reported at more micro levels, such as classrooms, the reliability or error associated with those results would need to be estimated as well. The goals of the measurement precision studies to be done here are to provide an estimate of the error inherent in any aggregate scores that are reported for the summative assessments and to judge the utility of the information given the estimates of error. It is possible that these studies will support the use of the summative assessment data for accountability purposes at some levels (e.g., districts) but not others (e.g., schools), because of the increased sampling error associated with smaller numbers of students.

Several methods have been proposed to estimate the reliability, or standard errors, associated with aggregate scores from statewide assessments. Yen (1997) used generalizability theory (G-theory) to estimate the reliability of school-level results for percent-above-cut statistics associated with the Maryland State Performance Assessment program and evaluated a criterion of achieving a standard error, of these percents, of 2.5% or less. She concluded that was an unrealistic criterion for performance assessments in a single subject area, but could be reached when evaluating a composite across subject areas. Her study illustrated the utility of G-theory for estimating standard errors for aggregate statistics, regardless of the item formats that are used.

Hill and DePascale (2003) asserted that the reliability of decisions at the school level should be evaluated from a decision consistency perspective. That is, if the assessment were repeated, would a school receive the same (AYP) classification? Hill and DePascale (2002) listed four methods for estimating school classification consistency. The first, "direct computation," is based on errors associated with each single classification and "uses areas under the normal curve to determine the probability of a correct classification" (p. 4). The second method is based on randomly dividing the students in a school into two groups and calculating the accountability statistics on each half. The third method involves randomly selecting (with replacement) multiple samples from a school, and the fourth method involves Monte Carlo simulation, where the parameters for a school are estimated and then random draws of students are made. In all four methods, the consistencies in schools' classifications are evaluated. Hill and DePascale recommend using at least two methods to offset the disadvantage of any single method.

Regardless of the method used to estimate the reliability of or error associated with aggregate summative assessment statistics used for accountability, it is important that the estimates address both measurement error and sampling error (Hill & DePascale, 2003; Linn, Baker, & Betebenner, 2002), as do the aforementioned approaches by Yen (1997) and Hill and DePascale (2002, 2003).

Simulation or empirical studies should also be conducted to evaluate the impact of factors outside of a school's control (or outside of the control of whatever the unit of inference is, such as a teacher) on the accountability results. For example, the inference made about a district or a school should not be statistically biased based on the number of students, the number of subgroups of students, or other factors beyond instruction. By estimating and using standard errors associated with aggregate scores when making accountability decisions, the validity of those decisions will be enhanced.

Simulation and other studies could also be used to inform accountability decisions such as how many years of data should be used to evaluate a district, school, or other unit of interest.

The degree to which derivative measures of summative assessment scores, such as "growth" measures, will be used in accountability systems is not known at the time of this writing. Any derivative measures would need to demonstrate evidence of reliability and validity. The Standards made this point when discussing what today might be considered a "growth" score: "When change or gain scores are used, the definition of such scores should be made explicit, and their technical qualities should be reported" (AERA et al., 1999, p. 167). Unfortunately, many of the current score derivatives, such as growth percentiles and value-added scores for teachers, have not been widely studied. As Brennan (2011) lamented, "to the best of my knowledge the subject of error variances and measures of precision for measures of growth is largely uncharted territory" (pp. 16–17).

**Validity Studies Based on Relations to Other Variables.** The use of summative assessment results for federal accountability purposes will certainly involve the use of achievement level results. In addition to the reliability studies previously mentioned, the previously mentioned studies supporting the use of achievement level standards are also relevant. However, additional studies are needed to support the utility of aggregate results based on achievement level results. For example, are the schools that are identified as not making adequate progress, based on percentages of "Proficient" or "on track" students, really the schools that should be flagged? Studies that could be designed to answer this question include using other measures of student achievement to classify schools into performance categories, and single-case studies where schools identified as over- or underperforming are carefully reviewed to evaluate the classification.

With respect to other measures of student achievement, at the high school level, changes in summative assessment scores for a school could be compared with the school's changes in scores on AP and college admissions tests. Perhaps student fees for these admissions tests could be paid for to remove the self-selection problem. At the middle school level, ACT's and the College Board's assessments for younger students (EXPLORE, PLAN, ReadiStep) could be used.

**Validity Studies Based on Testing Consequences.** The use of test scores for accountability has been accused of causing many problems, such as decreased teacher morale, increased pressure on students, and narrowing of the curriculum. As described earlier for purposes 1 through 3, these criticisms could be studied using comprehensive surveys of students and teachers, both before and after the implementation of the summative assessments. Surveys could be used to understand the effects on students (e.g., anxiety, educational aspirations), teachers (morale, retention, movement into non-tested subject areas, instruction), administrators (e.g., teacher recruitment and retention, effectiveness of school improvement), and parents (e.g., observations of their child, school choice). Teacher retention rates and teachers' movement into non-tested subject areas should also be tracked and studied.

### Summative Assessment Purpose 7:

> Provide valid, reliable, and fair information about students' achievement in ELA and mathematics that is equitable for *all students and subgroups of students.*

There are several features of the Smarter Balanced summative assessments that support equitable assessment across all groups of students. For example, the assessments are developed using the principles of universal test design; test accommodations are provided for students with disabilities; and Spanish-language versions of the math assessments will be developed. In addition, there is a specific work group for accessibility and accommodations, and the Consortium has developed seven sets of guidelines to facilitate accessibility of the assessments. These include general accessibility guidelines for item writing and reviewing (Measured Progress & ETS, 2012) and guidelines for creating audio, sign language, and tactile versions of the items. The Consortium also developed guidelines for item development that aim toward reducing construct-irrelevant language complexities

for English language learners (Young, Pitoniak, King, & Ayad, 2012), and comprehensive guidelines for bias and sensitivity (ETS, 2012b). These documents underscore the Consortium's commitment to fair and equitable assessment for all students, regardless of their sex, cultural heritage, disability status, native language, or other characteristics.

Irrespective of these proactive activities designed to promote equitable assessments, studies must be done to provide validity evidence that the assessments are fair for all groups of students. Many of the equity issues are delineated in the most recent version of the NCLB Peer Review Guidance (U.S. Department of Education, 2009b). For example, these guidelines recommend providing translations in appropriate languages and formats (p. 66), and they require statistical evidence of comparability across different language versions of assessments (p. 36). These guidelines also require that all students be included in the assessment, regardless of disability or English language proficiency status.

Of these requirements, statistical evidence of comparability across the English- and Spanish-language versions of the math assessments, and across standard and accommodated test administrations, is particularly important. For example, the Standards assert, "When multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability" (AERA et al., 1999, p. 99). Similarly, the ITC's Guidelines on Test Adaptation (Hambleton, 2005) state that "Test developers/publishers should apply appropriate statistical techniques to (a) establish the equivalence of the language versions of the test, and (b) identify problematic components or aspects of the test that may be inadequate in one or more of the intended populations" (p. 22). Thus, empirical analyses to evaluate the comparability of the English- and Spanish-language versions of the math summative assessments are needed. Similar evidence will be needed to evaluate the comparability of standard and accommodated tests.

To evaluate the degree to which the summative assessments are fulfilling the purpose of providing valid, reliable, and fair information that is equitable for all students, several studies are recommended. These studies are categorized here as validity evidence based on all five sources of evidence listed in the Standards.

**Validity Studies Based on Test Content.** Validity studies based on test content to support the equitability of the assessments will be based on the degree to which the planned universal test design, guidelines for assessing English language learners, and other fairness guidelines are implemented and followed. Documents regarding sensitivity review, and how items that were flagged for DIF were handled, should be reviewed. The test development processes and scoring processes are designed to minimize sources of construct-irrelevant variance that would inhibit fairness. The degree to which these procedures are followed and documented should be audited. Part of this audit should ascertain the degree to which students with disabilities, underrepresented minorities, and English language learners were included in the field tests, and the degree to which their special characteristics were addressed in scoring.

**Validity Studies Based on Internal Structure.** When evaluating the comparability of different variations of a test, such as different language versions of an assessment or accommodated test administrations, validity studies based on internal structure are most common (Sireci, Han, & Wells, 2008). These studies most often involve multi-group confirmatory factor analysis (CFA) (e.g., Ercikan & Koh, 2005). Weighted (multi-group) multidimensional scaling (MDS) has also been used for this purpose (e.g., Robin, Sireci, & Hambleton, 2003; Sireci & Wells, 2010). Both CFA and MDS involve simultaneous analysis of the dimensions underlying an assessment, and are used to assess whether the dimensionality is invariant across different versions of an exam. The CFA approach allows for statistical tests of different levels of invariance (number of dimensions, item factor loadings, correlations among factors, errors associated with factor loadings). The MDS approach does not typically involve statistical tests of invariance, but because it is exploratory, the dimensionality does not need to be modeled a priori.

Multi-group analyses of dimensionality can also be used to evaluate the comparability of scores for different subgroups of students who take the same test. For example, Day and Rounds (1998) used weighted MDS to look at structural invariance of an assessment across ethnic groups, and Marsh, Martin, and Jackson (2010) used multi-group CFA for this same purpose. The validity research agenda should use multi-group CFA or MDS to evaluate the invariance of test structure across diverse groups of students taking the standard versions of the summative assessments, as well as across students taking the standard and accommodated versions of the assessments.

In addition to comparing the dimensionality of the summative assessments across diverse groups of students, simpler analyses based on internal structure should also be performed. Essentially, these analyses involve breaking down the results of all studies of measurement precision to the subgroup level. Reliability estimates, conditional standard error functions, DC and DA estimates, and average standard errors should be reported for all subgroups and all different versions of the assessments. Given that reliability estimates are influenced by variability in students' responses, comparisons of measurement precision are better if based on estimates of the standard error of measurement.

One other important source of validity evidence to support equitable assessment for all is analysis of DIF across test variations and across subgroups of students. There are numerous procedures for evaluating items for DIF, and because excellent descriptions of these procedures exist (e.g., Clauser & Mazor, 1998; Holland & Wainer, 1993), they are not described here. DIF studies conducted for the summative assessments should include an effect size criterion to distinguish statistically significant DIF from substantively meaningful DIF (i.e., reflect construct-irrelevant variance). The presence of DIF does not necessarily indicate bias, and so DIF studies must be followed up by qualitative analysis to try to interpret the source of DIF. Finally, the DIF studies should evaluate the *aggregate* effect of DIF at the total test score level, or at least estimate how the presence of some DIF items may affect the typical test taker from a subgroup.

**Validity Studies Based on Response Processes.** The studies involving validity evidence based on response processes for purpose 1 are relevant here in that relevant subgroups of students should be included in those studies and the results should be broken down by subgroup. In particular, the amount of time that different groups of students take to respond to items, both with and without accommodations, should be studied. Any cognitive interviews or think-aloud protocols that are conducted to evaluate the skills measured by items should be inclusive in recruiting students. In addition, specific studies to evaluate accommodations for English language learners or students with disabilities should be conducted to determine whether the students are using the accommodations and find them helpful (e.g., Duncan et al., 2005).

**Validity Studies Based on Relations to Other Variables.** Two types of studies based on relations to other variables are relevant for validating that the summative assessments are equitable for all subgroups of students. The first are differential predictive validity studies that evaluate the consistency of the degree to which the assessments predict external criteria across subgroups of students. Zwick and Schlemer (2004) provide an excellent example of this type of analysis with respect to the differential predictive validity of the SAT across native English speakers and non-native English speakers. These studies will be particularly relevant for the "on track" and "college and career readiness" standards associated with the summative assessments. Of course, the caveats that were mentioned earlier regarding the validity of the external criteria apply here.

The second type of study involves a grouping variable as the external variable. Experimental studies that have looked at test accommodations fall into this category. For example, in some studies, students with and without disabilities are randomly assigned to test accommodation or standard test administration conditions. The validity hypothesis investigated is one of "differential boost," which states that students with disabilities will have larger score differences across the accommodated and standard conditions than students without disabilities, and that their scores will be higher in the accommodated condition (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000).

Non-experimental studies using grouping variables could also be conducted using an expected hypothesis of no difference across groups. For example, using changes in students' scale scores over time as the dependent variable, comparisons could be made across students of different ethnic groups, SES, sexes, and other demographic characteristics.

In addition to the studies previously described in this section, all other studies conducted on the general population could be broken down by subgroup to evaluate consistency of the results across subgroups, where sample sizes permit. For example, if multitrait-multimethod studies are conducted, a study of the invariance of results across subgroups may prove interesting.

**Validity Studies Based on Testing Consequences.** The analysis of the results from the summative assessments across subgroups of students will be a good starting point for understanding if there are differential consequences for certain types of students. In describing validity studies based on testing consequences for other purposes of the summative assessments, we discussed investigating the effects on instruction, teacher morale, and students' emotions and behaviors (e.g., dropout, course-taking patterns). These results should also be broken out by subgroup, but more importantly, the changes in instructional decisions for students should be investigated at the subgroup level. Important analysis questions include: Are minority students dropping out of school at higher rates than non-minorities? Are the success rates for remedial programs higher for certain types of students?

# VI. Validity Agenda for Interim Assessments

The Smarter Balanced interim assessments differ from the summative assessments in that they are optional, include both secure and non-secure components, are customizable across users, can be administered multiple times within a school year, and are designed to provide information at a finer level of detail with respect to students' strengths and weaknesses in relation to the CCSS. The validity studies described for the summative assessments are essentially all relevant to the interim assessments, but additional validation work needs to address the degree to which the interim assessments provide the intended diagnostic information and are useful to teachers, administrators, and other educators for improving instruction and student learning.

As indicated in Chapter III, four purpose statements for validation are associated with the interim assessments. The proposed studies to support the validity of these statements are described in this section.

### Interim Assessment Purpose 1:

> Provide valid, reliable, and fair information about students' progress toward mastery of the skills measured in ELA and mathematics by the summative assessments.

To support this purpose, validity evidence should confirm that the knowledge and skills being measured by the interim assessments cover the knowledge and skills measured on the summative assessments and that the interim assessment scores are on the same scale as those from the summative assessments. As indicated in Table 2 (p. 15), the studies providing this evidence will primarily be based on test content, internal structure, and response processes.

**Validity Studies Based on Test Content.** The content validity studies described for the summative assessments will gather data relevant to the interim assessments. However, an additional level of analysis will be required to support the validity of reporting students' performance at the content cluster levels. The sample results of a summary of a content validity study that were reported in Figure 4 (p. 30)suggest how results could be summarized for the content clusters targeted by the interim assessments. Moreover, the data from such studies could be used to select the best items for interim assessment purposes. That is, items that are rated as measuring their intended CCSS "very well" could be selected for the interim assessment item bank.

The interim assessments are intended to help teachers focus assessment on the most relevant aspects of their instruction at a particular point in time. Thus, the interim assessments should better align with teachers' instruction, if the content clusters are appropriately selected. To evaluate this intended benefit of the interim assessments, surveys could be given to teachers regarding the instructional objectives that they cover at several points during the school year (i.e., scope and sequence survey). Then, the content clusters that were administered to these teachers' students at specific points in time can be evaluated ex post facto, and the match between what was taught and what was assessed can be calculated. This type of survey could be coupled with survey questions regarding the utility of the interim assessments, which is relevant to purpose 2.

**Validity Studies Based on Internal Structure.** Scores from the comprehensive interim assessments are intended to be on the same scale as those from the summative assessments, to best measure students' progress toward mastery of the knowledge and skills measured on those assessments. This intent requires linking the scores from the interim and summative assessments. Given that many of the items in the interim assessment item bank will also be used on the summative assessments, it is assumed that some type of common item equating will be used to place students' performance on the interim assessments on the summative assessment score scale. This equating should be evaluated to support the inferences about how well students are likely to do on the summative assessments based on their interim assessment scores. Studies in this area would

include an audit of the equating procedures, such as analysis of equating error and analysis of DIF of equating items across groups of students defined by state, ethnicity, or other factors (or a more formal population invariance study; Dorans, 2004). In addition, the degree to which interim assessment items fit the IRT models determined by the summative assessment scale should be ascertained. The fit of the equating items to this model will be of particular interest.

Also under the realm of internal structure is evidence regarding the reliability or measurement precision of scores from the interim assessments. Less measurement precision relative to that of the summative assessments is tolerable because (a) the stakes are lower, (b) there will be multiple assessments, and (c) these assessments supplement the summative assessments, on which higher-stakes decisions are based. However, studies should be conducted to ascertain the reliabilities and errors of measurement associated with any scores reported from the interim assessments so that they can be properly interpreted. If achievement level classifications are made on the basis of these assessments, then estimates of DC and DA should also be calculated.

Studies should also be conducted to evaluate the quality and accuracy of local scoring of the performance tasks associated with the interim assessments. Having trained scorers rescore samples of locally scored tasks, and the degree to which local scorers can assign similar scores to training sets of responses, will provide evidence regarding the quality of local scoring.

**Validity Studies Based on Response Processes.** Interim Assessment Purpose 1 relates to skills measured on the summative assessments, and so the validity studies based on response processes that were described for the summative assessments are relevant here in order to confirm that the items are measuring higher-order skills. The response process studies for Summative Assessment Purpose 1 should include items that will be used on the interim assessment. The results from these studies should be used to "assure that each item or task clearly elicits student responses that support the relevant evidence statements and thus are aligned to the associated claims and standards" (ETS, 2012c, p. 4).

**Interim Assessment Purpose 2:**

> Provide valid, reliable, and fair information about students' performance at the content cluster level, so that teachers and administrators can track student progress throughout the year and adjust instruction accordingly

As shown in Table 2, validity evidence to support this purpose of the interim assessments will rely on studies of test content, internal structure, and testing consequences.

**Validity Studies Based on Test Content.** Assuming that the content validity/alignment studies described for the summative assessments are conducted, all items on those assessments will be rated regarding the degree to which they measure their intended CCSS and their intended cognitive skills. These studies should be extended to include the items on the interim assessments that do not overlap with the summative assessments. However, an additional study is needed to support purpose 2. A study should be conducted to confirm that the content clusters associated with the interim assessments represent helpful groupings of CCSS that are useful for tracking progress and adjusting instruction. These studies would evaluate whether the specific groupings of standards from the CCSS into content clusters is instructionally beneficial.

Like all content validity studies, this study would require SMEs. Rather than reviewing items, the SMEs would review the CCSS that were used to create the content clusters for each claim area. Their task could be to group the standards in a way that would be best for providing instructionally relevant information. Their groupings of standards could then be compared to how the standards were grouped into the content clusters, and the consistency across the actual and SME-derived clusters could be calculated. Alternatively, the SMEs could review the content clusters and rate them

for their instructional relevance, and make comments about whether and how they might be rearranged.

**Validity Studies Based on Internal Structure.** Information regarding the reliability and measurement error of cluster-level score reporting should be provided. In addition, the degree to which different clusters are correlated should also be reported, to see if clusters measuring different assessment targets or claims correlated less than clusters measuring the same claims and targets. A multitrait-multimethod approach could be used, using the different item formats and different claim areas as methods and traits, respectively (Pitoniak, Sireci, & Luecht, 2002).

**Validity Studies Based on Testing Consequences.** The interim assessments are designed to "provide more immediately actionable data for teachers and students" (ETS, 2012c). A primary validity question to be studied is: Do the content cluster results help teachers and administrators track student progress and adjust instruction? To assess the effects on instruction, studies should be conducted to (a) track the use of the interim assessments and their associated supports (e.g., user tutorials), (b) assess the degree to which teachers and administrators find the system easy to navigate, and (c) assess the degree to which teachers and administrators value the information provided and use it to adjust instruction. Studies could also be conducted to ascertain students' impressions of the system.

Tracking the use of the interim assessments should be straightforward, assuming that most of the assessments are accessed online and that these testing occasions are captured by the system. Procedures should be in place to track any uses that are not online. Surveys of teachers and administrators will be needed in order to understand the degree to which these educators find the system useful and easy to navigate. Surveys of teachers and administrators will also be needed to ascertain the effects on instruction. As part of that study, "high use" teachers and schools should be identified and selected for further inquiry. Surveys, interviews, and focus groups of these teachers should be conducted, to learn about how they used interim assessment results to improve instruction.

### Interim Assessment Purpose 3:

> Provide valid, reliable, and fair information about individual and group (e.g., school, district) performance at the claim level in ELA and mathematics, to determine whether teaching and learning are on target.

As shown in Table 2, validity evidence to support this purpose of the interim assessments will rely on studies of internal structure, relations to other variables, and testing consequences.

**Validity Studies Based on Internal Structure.** This purpose statement is similar to purpose 2, with the difference being that rather than a focus at the content cluster level, the focus here is on the claim level. The studies described for purpose 2 are all relevant here. The additional studies needed would need to evaluate the reliability and precision of the claim scores at the group level. It is assumed that claim-level information will be provided by the interim assessments during the school year, and so estimates of the precision of this information should be provided, using the same types of internal structure studies described for purposes 1 and 2.

**Validity Studies Based on Relations to Other Variables.** Given that the interim assessments will provide information at the claim level throughout the school year, it would be good to study the degree to which the information provided for individual students or groups of students is consistent with other measures of their performance relative to the CCSS. One way to study this relationship is to see how well the claim scores for the interim assessments predict claim scores on the summative assessments. In particular, it would be interesting to assess the degree to which students who are considered "on target" or "not on target" are classified similarly on the summative assessments. More interesting, however, would be to qualitatively study students who are mispredicted. That is, if

a student did poorly on an interim assessment but well on a summative assessment, is that a success story or a story of poor measurement by the interim assessment? If other measures of student achievement are available, they would be helpful for shedding light on this issue, but it may be difficult to find other measures tied to the same CCSS that specific interim assessments are measuring. Nevertheless, assessments such as NWEA's Measures of Academic Progress or Curriculum Associates' iReady assessment may be relevant.

**Validity Studies Based on Testing Consequences.** As mentioned for purpose 2, the intended consequence of the interim assessments is to connect the assessments to instruction to improve student learning. The validity studies based on testing consequences that were described for purpose 2 are all relevant here, with the only difference being that the information provided would be at the claim level and would be extended to groups of students. Therefore, the studies described earlier should include these factors to provide validity evidence in support of purpose 3. In addition, should in-class activities (classroom interaction tasks) become part of the interim assessment system, their effectiveness should be a focus of the surveys, interviews, and focus groups associated with the studies mentioned earlier.

**Interim Assessment Purpose 4:**

> Provide valid, reliable, and fair information about student progress toward the mastery of skills measured in ELA and mathematics *across all students and subgroups of students.*

Validity evidence in support of this purpose should come from all five sources. The validity studies based on test content that were described with respect to purposes 1 and 2 provide the starting point for equitable measurement across all students. The validity studies based on internal structure should report any estimates of reliability, measurement precision, DC, or DA separately for all subgroups of students, and for students who take different variations of the interim assessments. In addition, it should be documented that access to the interim assessments has been provided to all students, as was discussed in relation to the summative assessments. Such access should include appropriate test accommodations for students with disabilities and English language learners.

The Peer Review Guidance for NCLB assessments stipulates that states should "Provide written documentation of criteria for local assessments, which ensures technical quality and comparability to state assessments of locally used tests for ALL subgroups and content areas (includes modified/alternate assessments)" (U.S. Department of Education, 2009b, p. 32). The interim assessment system allows states and districts to create their own assessments from the banks of items, and so the technical quality of these local assessments will need to be studied to ensure that they provide comparable measurement across all groups of students.

# VII. Research Agenda for Formative Assessment Resources

The third component of the Smarter Balanced Assessment Consortium is *formative tools and processes,* referred to in this report as *formative assessment resources*. These resources are not assessments per se, and so their evaluation does not neatly fit into the Standards' five sources of validity evidence. Rather, these resources are intended to work with the summative and interim assessments to increase their utility for improving instruction and helping students learn. Essentially, the formative assessment resources are what puts the "balance" in the Smarter Balanced Assessment Consortium.

The purposes of the formative assessment resources that are the focus of the comprehensive research agenda were listed in Chapter III, and, for convenience, are repeated here.

The purposes of the Smarter Balanced *formative assessment resources* are to provide measurement tools and resources to:

1. Improve teaching and learning.

2. Monitor student progress throughout the school year.

3. Help teachers and other educators align instruction, curricula, and assessment.

4. Help teachers and other educators use the summative and interim assessments to improve instruction at the individual student and classroom levels.

5. Illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning.

To accomplish these goals, the formative assessment resources will provide tools and professional development materials including a "Digital Library," learning modules (lesson plans, templates, curriculum resources, evidence collection tools, video clips of classroom instruction and teacher analysis, descriptive feedback strategies, follow-up planning materials), online assessment literacy training products, webinars, tutorials, and PowerPoint presentations. To oversee the development, implementation, and maintenance of these resources, extensive collaboratives will be established, including:

- National Advisory Panel

- Digital Library Review Board

- State Leadership Teams

- State Networks of Educators

- Formative Assessment Practices and Professional Learning Work Group

The research agenda for this component of the Consortium will be an evaluation of the products developed for these purposes and of the processes for developing them. Studies comprising this evaluation should involve (a) confirming the development and successful implementation of all planned formative assessment resources; (b) evaluating usage statistics of all tools and other resources; (c) review of all documents supporting the system; (d) comprehensive surveys of the collaborative leadership involved in overseeing the products and processes; (e) comprehensive surveys of users of the resources (teachers, administrators, students, parents); and (f) case studies of teachers and administrators who are frequent users of the resources. It should also be confirmed that teachers were involved in the development and review of these materials.

### Confirming Development and Successful Implementation of Products

The RFP for the "Digital Library with Formative Assessment Practices and Professional Learning Resources for Educators," hereafter referred to as RFP-23, specifies the development of several products using specific processes. An important step in the evaluation of the formative assessment resources is to confirm that all of the deliverables associated with this contract were satisfied. For example, RFP-23 calls for the development of at least 50 exemplar instructional modules (p. 26). The successful creation of these modules, and other tasks, will be audited as part of the evaluation. In addition, goals related to the review and implementation of all resources will be reviewed in this evaluation. This step will merely confirm that the intended products and activities occurred and note the timeliness of the deliverables. The quality of the products and their implementation will be evaluated using other activities described later in this chapter.

### Evaluating Usage Statistics

The formative assessment resources are designed to be used by teachers, administrators, and even parents and students. If these resources are not understood and found useful, the system will be unbalanced, which will inhibit the goals of the entire Consortium. One way to evaluate the utility of the resources is to analyze their usage statistics. RFP-23 specifies reporting monthly usage statistics (p. 71). These statistics should be analyzed over time. Formative evaluation should inform the Smarter Balanced leadership about which resources are being used and which are not, so that better advertising or improvement of the underutilized resources can be considered. Analysis of usage data should be broken down by state, and by important subcategories within states, such as type of school, geographic region, percentage of certain subgroups of students within a school (English language learners, low-SES, etc.), and, where possible, demographics of the users.

### Document Review

RFP-23 specifies several documents that are important to the integrity of the formative assessment resources. These documents include:

- Comprehensive development strategy
- Biannual implementation reports
- Documentation of component plans and processes
- Description of recruiting and creation of leadership committees (State Leadership Teams, State Networks of Educators)
- Records of decision-making by leadership committees
- Technical documentation of system components

These documents will be reviewed to ensure that products are developed as intended and processes are followed. Any problems discovered in the documents should be followed up on to see if they were properly resolved. In addition, RFP-23 requires the contractor to perform and document quality assurance testing (pp. 69–70). This documentation will also be reviewed as part of the evaluation. Monitoring reports on user comments (p. 71) will also be reviewed and reported on.

### Surveys, Interviews, and Focus Groups of Leadership

The plan for developing, implementing, and improving the formative assessment resources calls for full participation of educators throughout the Consortium. In particular, the State Networks of Educators will involve carefully selected end-users of the resources. In the evaluation, the five aforementioned collaboratives of leaders (National Advisory Panel, Digital Library Review Board, State Leadership Teams, State Networks of Educators, Formative Assessment Practices and

Professional Learning Group) will be solicited to participate in surveys, interviews, or focus groups to obtain their impressions of the process, the quality of the products, and the degree to which the formative assessment resources are accomplishing the intended goals. In addition, the intended representation of the membership of these committees with respect to geographic region, subject expertise, representation of special populations, and other characteristics will be evaluated.

## Surveys of Users

The evaluation activities previously described will provide information on the quality of the products and processes and the degree to which users are accessing the resources. However, it is also critical to gather information regarding the degree to which the resources are perceived as being helpful to educators. RFP-23 includes the development of a survey to assess the effectiveness of the regional meetings (p. 23). The results from that survey should be considered in the evaluation. More importantly, however, we recommend that the research agenda include large-scale surveys of all users. Given that the bulk of the resources must be accessed online, *we recommend that user surveys be implemented as part of the system*. That is, at strategic points in time, users should be required, or heavily encouraged, to take brief surveys, for the Consortium to obtain their opinions regarding the usefulness of the materials and how they use the resources in their instructional practices. The surveys should target the specific aspects of the resources (e.g., lesson plans, evidence collection tools, assessment literacy training products, understanding how to use summative and interim data to improve instruction, etc.). Surveys to evaluate training programs delivered as part of the implementation of the resources (e.g, RFP-23, p. 65) are also needed. These surveys are needed in order to provide evidence that the formative assessment resources are having an impact on classroom practices.

Teacher survey data could also be used to create an implementation index for participating teachers, and those data could be correlated with students' test scores. In particular, it would be interesting to correlate teachers' implementation data with the progress that students make *within the school year* while they have the teacher. If all aspects of the system work as intended, teachers who successfully use the formative assessment resources will be able to use the summative and interim assessment results to improve instruction, and will see greater gains for their students, relative to comparable teachers who do not use the resources.

It is also important to gather data on the degree to which parents, students, teachers, administrators, and others understand the reports from the summative and interim assessments. These data can be gathered using surveys to obtain opinions of the reports, and also by testing these individuals regarding the accuracy of their interpretations (Wainer et al., 1999).

## Case Studies of Frequent Users

The usage data for the formative assessment resources can be used to identify teachers and administrators who are frequent users. A sample of these frequent users can be selected and recruited for in-depth study of how they use the resources. The appropriateness of their practices can be documented, and ideas for improving the resources, and for sharing the lessons learned by these teachers and administrators, can be reported.

# VIII. Summary: The Smarter Balanced Assessment Consortium Validity Argument

The preceding chapters describe a multitude of studies that comprise the comprehensive research agenda for the Smarter Balanced Assessment Consortium. The presentation of the agenda according to the different components of the system may result in two misleading perceptions. These potential misleading perceptions are:

- The research agenda is too ideal to be practical because the agenda is too voluminous and optimistic.
- The research agenda is fragmented and so does not address the holistic goals of the Consortium.

In this chapter, we put those potential misperceptions to rest by illustrating the integration of studies across the various components and illustrating how many of the studies are already addressed in the test development and formative assessment resources development activities.

The integration of the various studies results in an agenda that, if properly implemented, can provide a convincing validity argument to support the goals of the Consortium as stated in its Theory of Action (Appendix A). Bennett (2010) posited six questions that should be posed to evaluate a theory of action for a comprehensive assessment system such as Smarter Balanced. These seven questions are:

- Is the theory of action logical, coherent, and scientifically defensible?
- Was the assessment system implemented as designed?
- Were the interpretive claims empirically supported?
- Were the intended effects on individuals and institutions achieved, and did the postulated mechanisms appear to cause those effects?
- What important unintended effects appear to have occurred? (p. 82)

The first question can be addressed by a thoughtful review of the Smarter Balanced Theory of Action as a preliminary step in the evaluation. Our impression is that the theory is defensible, which is supported by the fact that we were able to create a comprehensive research agenda to address its goals. The second question can be answered by analysis of the results from the studies outlined in this report, specifically the audit studies listed in Chapters III and VII and the studies regarding validity evidence based on testing consequences that involve surveys, interviews, and focus groups of stakeholders (described in Chapters IV through VII).

What most people think about when considering validation of an assessment system are the third and fourth questions posed by Bennett (2010). We, and many others (e.g., Haertel, 1999; Messick, 1989; Shepard, 1993), would also include the sixth question. These three questions require validity evidence beyond typical test development activities, and require evidence stemming from all five sources stipulated in the Standards. It is around these three questions that the majority of studies described in Chapters V through VII are centered.

The Smarter Balanced Theory of Action is based on seven principles (Smarter Balanced, 2010). These principles are presented in Appendix A and are presented here in more abbreviated form:

1. Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of an integrated system.
2. Assessments produce evidence of student performance.
3. Teachers are integrally involved in the development and scoring of assessments.
4. The development and implementation of the assessment system is a state-led effort with a transparent and inclusive governance structure.

5. Assessments are structured to continuously improve teaching and learning.
6. Assessment, reporting, and accountability systems provide *useful information on multiple measures* that is educative for all stakeholders.
7. Design and implementation strategies adhere to established professional standards. (pp. 32–33)

A review of the purpose statements on which this comprehensive research agenda is based (see Chapter III) makes clear that the agenda is focused on evaluating the degree to which these principles are realized. To pull the comprehensive research together—that is, to document the validity argument for Smarter Balanced in a coherent manner to best inform stakeholders and the general public—a report should be produced that indicates how the various pieces of evidence gathered through the research agenda confirm that these seven principles are realized. If the research agenda outlined in this report is followed, it will provide ample evidence that could be organized in a reader-friendly report that is organized around these seven principles. It is clear that the research agenda outlined here addresses the seventh principle. Our review of Smarter Balanced activities to date supports the fourth principle, and evidence for the collaboration could easily be documented. The remaining five principles would be supported by evidence from the studies described in this report.

## Summarizing the Validity Evidence

As promised earlier in this chapter, the validity studies described in this report will appear less daunting when the overlap of studies across the different purposes and components of the Smarter Balanced assessment system is accounted for. This integration is presented in Tables 7 and 8. Table 7 presents brief descriptions of each proposed study in the form of short labels, indicates the purposes that each study addresses, and provides a unique number for each study. It also lists the page numbers in this document that refer to each study. Table 8 uses this numbering system to illustrate the places where such studies are already accounted for in current or planned Smarter Balanced activities. Table 8 is also available as an Excel file, so that its data can be sorted by columns to facilitate different research planning activities. It may be tempting to prioritize the studies based on the number of check marks in each row of Table 7, but because the purposes in the columns are not equal in importance, and because the contribution of each study to the validity argument will not be equal, such an interpretation would be an oversimplification.

Table 7. Listing of Studies by Source of Evidence and Testing Purpose.

| Study Number and Description | Page Numbers | Evidence Sources | Summative Assessment Purpose | | | | | | | Interim Assessment Purpose | | | | Formative Resources Purpose | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 |
| 1 Audit of test construction practices | 16–18 | 1, 3 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | | √ | √ | √ |
| 2 Analysis of measurement precision | 17–18, 34, 36–37, 51–52, 53, 56, 58–59 | 3 | | √ | √ | √ | √ | √ | √ | | | | | | | | | |
| 3 Audit of test administration | 17, 18 | 1, 5 | √ | | | | | | √ | | | √ | | | √ | | | |
| 4 Evaluation of scoring | 17, 19 | 1, 3 | √ | √ | √ | √ | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 5 Analysis of scaling and equating | 17, 19, 46–48, 58–59 | 3 | | √ | √ | | | √ | √ | √ | | | | | | | | |
| 6 Evaluation of standard setting | 17, 19–21, 36–45 | 1, 3, 4 | √ | √ | √ | √ | √ | √ | √ | | | √ | | √ | | | √ | |
| 7 Evaluation of fairness | 17, 22, 52, 59 | 1–5 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 8 Evaluation of equitable particip. & access | 17, 23–24, 52, 54, 59 | 1, 5 | √ | | | | | | √ | | | √ | | | | | √ | √ |
| 9 Audit of test security | 17, 24–25, 48 | 3, 4 | √ | √ | √ | | | √ | | √ | | | | | | | | |
| 10 Content validity and alignment | 25–31, 36, 39–40, 46, 53, 56–58, 86 | 1 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 11 Evaluating ECD | 25, 31 | 1, 2 | √ | √ | √ | √ | | | | √ | √ | √ | | | | | | |
| 12 IRT residual analysis | 31–34, 46 | 3 | | √ | | √ | | √ | √ | √ | | | | | | | | |
| 13 Reliability and standard error estimation | 17–19, 31–34, 36–37, 50–52, 53–54, 56–59 | 3 | √ | √ | √ | √ | √ | √ | √ | √ | √ | | | | | | | |
| 14 Cognitive skills and item response time | 24, 35, 54, 56 | 2 | √ | | | √ | | √ | | √ | | | | | | | | |

| Study Number and Description | Page Numbers | Evidence Sources | Summative Assessment Purpose | | | | | | | Interim Assessment Purpose | | | | Formative Resources Purpose | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 |
| 15 Cognitive interviews, think-aloud | 35, 54, 56 | 2 | √ | | | | √ | | √ | | √ | | | | | | | |
| 16 Decision consistency and accuracy | 36–38,41, 56–57, 59 | 3 | | √ | √ | √ | | √ | √ | √ | | √ | | | | | | |
| 17 Cut-score standard errors | 36–37 | 3 | | √ | √ | √ | | √ | √ | √ | | √ | | | | | | |
| 18 Criterion-related validation of "on track" | 37–38 | 4 | | √ | | | | | | | | | | | | | | |
| 19 Educator interviews, focus groups, surveys | 38–39, 44–45, 49–50, 58–59 | 5 | | √ | √ | | | | | | | | | | | | | |
| 20 Criterion-related validation of readiness | 39–45 | 4 | | √ | √ | √ | | √ | | √ | | | | | | | | |
| 21 Surveys of postsecondary educators | 45, 62 | 5 | | | √ | √ | | | | | | | | | | | | |
| 22 Analysis of enrollment, dropout, courses | 38, 45, 55 | 5 | | | | | √ | | √ | | | | √ | | | | | |
| 23 Teacher morale surveys | 45, 52, 55 | 5 | | | | | | √ | √ | √ | √ | √ | √ | √ | | √ | √ | √ |
| 24 Teacher surveys on changes in students | 45, 49, 52–53, 55, 59, 61–62 | 5 | | √ | √ | √ | √ | | √ | √ | √ | √ | √ | √ | √ | | | √ |
| 25 Student morale and aspirations surveys | 45, 52 | 5 | | | √ | | √ | | | | | | | | | | | |
| 26 Evaluation of vertical scale | 46–48 | 3 | | √ | √ | √ | | √ | √ | √ | | | | | | | | |
| 27 Criterion-related studies re: gain/growth | 49, 52, 59 | 4 | | √ | | √ | | √ | √ | √ | √ | √ | | | | | | |
| 28 Follow-up on specific student decisions | 49–50 | 5 | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | √ | √ |
| 29 Sensitivity to instruction | 49–50 | 4 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | | | | |

| Study Number and Description | Page Numbers | Evidence Sources | Summative Assessment Purpose | | | | | | | Interim Assessment Purpose | | | | Formative Resources Purpose | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 |
| 30 Analysis of classroom artifacts | 49–50 | 4, 5 | | | | | √ | | | | √ | | | √ | √ | √ | √ | √ |
| 31 Score report utility and clarity | 31, 49–50, 61–62 | 5 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 32 Analysis of report usage rates | 49–50 | 5 | | | | | √ | | | | | | | √ | √ | √ | √ | √ |
| 33 Analysis of reliability of aggregate stats | 50–52, 59 | 3 | √ | | | | √ | √ | | | | √ | | √ | | | | |
| 34 Generalizability studies | 17–20, 51 | 3 | | √ | √ | √ | | √ | | | | | | | | | | |
| 35 Item parameter drift | 48 | 3 | | √ | √ | √ | | | | √ | | | | | | | | |
| 36 Audit of UTD and sensitivity review | 54 | 1 | √ | | | | | √ | √ | | | | √ | | | | | |
| 37 Audit of test accommodations | 53 | 1, 5 | √ | | | | | √ | √ | | | | √ | | | | | |
| 38 Differential item functioning | 17, 18, 21–22, 53–54, 56–57 | 1, 3 | √ | | | | | √ | √ | | | | √ | | | | | |
| 39 Differential predictive validity | 54 | 4 | √ | | | | | √ | √ | | | | √ | | | | | |
| 40 Invariance of test structure | 19, 53–54, 56–57 | 3 | √ | | | | | √ | √ | | | | √ | | | | | |
| 41 Analysis of group differences | 54–55 | 4 | | √ | √ | √ | | √ | √ | | | | √ | | | | | |
| 42 Multitrait-multimethod | 55, 57 | 3, 4 | √ | | | | | | √ | | √ | √ | √ | | | | | |
| 43 Scope and sequence curriculum survey | 56–57 | 1, 5 | | | | | √ | | | | √ | √ | | √ | √ | √ | √ | √ |
| 44 Validation of content clusters | 56 | 1, 3 | | | | | | | | | √ | √ | | | | | | |
| 45 Analysis of interim usage statistics | 57–58 | 5 | | | | | | | | √ | √ | √ | √ | | | | | |

| Study Number and Description | Page Numbers | Evidence Sources | Summative Assessment Purpose | | | | | | | Interim Assessment Purpose | | | | Formative Resources Purpose | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 |
| 46 Surveys, interviews, focus groups of (high) users of interim assessments | 57–58 | 5 | | | | | | | | √ | √ | √ | √ | | | | | |
| 47 Audit of formative resources development and implementation | 61–62 | 1, 5 | | | | | | | | | | | | √ | √ | √ | √ | √ |
| 48 Analysis of usage stats for formative | 61–62 | 5 | | | | | | | | | | | | √ | √ | √ | √ | √ |
| 49 Surveys of collaborative leadership | 61–62 | 5 | | | | | | | | | | | | √ | | √ | √ | √ |
| 50 Educator formative assessment surveys | 61–63 | 5 | | | | | | | | | | | | √ | √ | √ | √ | √ |
| 51 Formative assessment user surveys | 62 | 5 | | | | | | | | | | | | √ | √ | √ | √ | √ |
| 52 Parent, student formative surveys | 48–49, 63 | 5 | | | | | | | | | | | | √ | | | | √ |
| 53 Case studies of frequent users | 62 | 5 | | | | | | | | | | | | √ | √ | √ | √ | √ |
| 54 Critique of Theory of Action | 63–64 | 5 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 55 Summary of validity evidence acc. to 7 principles | 64–68 | 5 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |

Note: Evidence Sources: 1 = Test Content, 2 = Response Processes, 3 = Internal Structure, 4 = Relations to Other Variables, 5 = Testing Consequences

Table 8 (to be populated): Connecting Recommended Studies to Current Activities and RFPs

| Study and Number | Source of Evidence | Contract | Summative Assessments | Interim Assessments | Formative Assessment Resources |
|---|---|---|---|---|---|
| 1 TC audit | | | | | |
| 2 Meas. precision | | | | | |
| 3 Administration audit | | | | | |
| 4 Evaluation of scoring | | | | | |
| 5 Scaling and equating | | | | | |
| 6 Standard setting | | | | | |
| 7 Evaluation of fairness | | | | | |
| 8 Equity | | | | | |
| 9 Audit of test security | | | | | |
| 10 Content validity | | | | | |
| 11 Evaluating ECD | | | | | |
| 12 IRT residual analysis | | | | | |
| 13 Reliability and SE | | | | | |
| 14 item response time | | | | | |
| 15 Cognitive interviews | | | | | |
| 16 DC, DA | | | | | |
| 17 Cut-score SE | | | | | |
| 18 Criterion-related OT | | | | | |
| 19 Educator surveys | | | | | |
| 20 Readiness | | | | | |
| 21 Postsecondary surveys | | | | | |
| 22 Dropout | | | | | |
| 23 Teacher morale | | | | | |
| 24 Change surveys | | | | | |

| Study and Number | Source of Evidence | Contract | Summative Assessments | Interim Assessments | Formative Assessment Resources |
|---|---|---|---|---|---|
| 25 Student morale | | | | | |
| 26 Vertical scale | | | | | |
| 27 Gain (growth) | | | | | |
| 28 Student decisions | | | | | |
| 29 Sensitivity | | | | | |
| 30 Classroom artifacts | | | | | |
| 31 Score reports | | | | | |
| 32 Report usage rates | | | | | |
| 33 Aggregate stats | | | | | |
| 34 G-studies | | | | | |
| 35 Item parameter drift | | | | | |
| 36 UTD and sensitivity | | | | | |
| 37 Test accommodations | | | | | |
| 38 DIF | | | | | |
| 39 Diff. prediction | | | | | |
| 40 Invariance | | | | | |
| 41 Group differences | | | | | |
| 42 MTMM | | | | | |
| 43 Scope and sequence | | | | | |
| 44 Content clusters | | | | | |
| 45 Interim usage | | | | | |
| 46 Surveys high users | | | | | |
| 47 Formative audit | | | | | |
| 48 Formative usage | | | | | |
| 49 Collabor. leadership | | | | | |

| Study and Number | Source of Evidence | Contract | Summative Assessments | Interim Assessments | Formative Assessment Resources |
|---|---|---|---|---|---|
| 50 Educator FA surveys | | | | | |
| 51 FA user surveys | | | | | |
| 52 Parent/student surveys | | | | | |
| 53 Case studies: users | | | | | |
| 54 Theory of Action | | | | | |
| 55 Summary of validity | | | | | |

# IX. Ongoing Validation Activities and Support Systems

Validation can be thought of as a great job for a masochist because, in a sense, one can never absolutely "prove" that an assessment is totally valid for the complex purposes to which it is put (Haertel, 1999), and because assessments are dynamic, and they, and the populations that they assess, change over time, validation is an ongoing, essentially perpetual, endeavor. Nonetheless, at some point, decisions must be made regarding whether sufficient evidence exists to justify the use of a test for a particular purpose. Most of this report has focused on the purpose of conducting studies to provide such evidence and documenting the evidence into a coherent validity argument that would satisfy professional testing standards, federal peer review, and legal challenges. However, our professional responsibilities also require us to think toward the future, beyond the current funding for Smarter Balanced, and consider the potential positive and negative consequences that should be addressed in longer-range validation studies.

At this juncture, a few potential validity activities appear in the crystal ball. One is studying the degree to which products and processes provided by the Consortium persevere and are used over time. The Consortium's processes, products, and activities are designed to produce an enduring collaboration and resources that should outlive the Consortium. Thus, studying the long-term effects of Smarter Balanced on instruction, within and outside the Consortium states, would be an interesting research area.

Another area of interest is the specific uses of the Smarter Balanced assessments and formative resources beyond the currently anticipated uses. It is quite possible that states, districts, and schools will use the assessments for purposes that they think are useful and valid, but that are not currently anticipated. Some of these uses may be appropriate and creative; others may be problematic or even damaging. States and districts will certainly use some assessments and tools for educator accountability, and so the validity of such use is an area in need of future research.

Although all important areas of future research cannot be anticipated at this time, it is still wise to consider the support systems that Smarter Balanced can put in place to facilitate future validity research. For example, other large-scale assessment programs, such as NAEP, TIMSS, and PISA, make data available for secondary analyses. Occasionally, these programs provide grant money to support such secondary analyses. The types of studies to be funded can be specified in advance, or, preferably, applicants for funding could be asked to submit their own ideas for research to study what they believe are important validity questions.

Another example of a support system is the College Board's "validity research study service." This service is essentially a data-sharing agreement between the College Board and postsecondary institutions, whereby the institutions can send course grade information to the College Board and it will match the data with SAT scores and other College Board assessment scores. These matched data sets can then be used to conduct local validity studies for each institution.

In considering potential validity studies that will be important in the future, and by establishing research support systems, validity research for Smarter Balanced can outlive the formal research studies that will comprise the documented validity argument for the Consortium.

# References

Achieve, Inc. (2006). *An alignment analysis of Washington state's college readiness mathematics standards with various local placement tests*. Cambridge, MA: Author.

ACT. (2005a). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade 12 mathematics: Process report*. Iowa City, IA: Author.

ACT. (2005b). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade 12 mathematics: Special studies report*. Iowa City, IA: Author.

ACT. (2005c). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade 12 mathematics: Technical report*. Iowa City, IA: Author.

ACT. (2006). *Ready for college, ready for work. Same or different?* Iowa City, IA: Author.

ACT. (2010). *Issues in college readiness: What are ACT's college readiness benchmarks?* Iowa City, IA: Author.

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*, 955–959.

Allen, J., & Sconing, J. (2005). *Using ACT assessment scores to set benchmarks for college readiness* (ACT Research Report Series 2005-3). Iowa City, IA: ACT.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Evaluation Association. (2004). *Guiding principles for evaluators*. Retrieved from http://www.eval.org/publications/guidingprinciples.asp

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*, 70–91.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21–29.

Bock, R. D., Gibbons, R. D., & Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement, 12*, 261–280.

Brennan, R. L. (2002). *Estimated standard error of a mean when there are only two observations* (CASMA Technical Note Number 1). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, University of Iowa.

Brennan, R. L. (2011). *Using generalizabilty theory to address reliability issues for PARCC assessments: A white paper*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Briggs, D. C. (2012, April). *Making inferences about growth and value-added: Design issues for the PARCC consortium.* Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Briggs, D. C. (2013). Measuring growth with vertical scales. *Journal of Educational Measurement, 50*(2), 204–226.

Camara, W. (2012, April). *Defining and measuring college and career readiness: Developing performance level descriptors and defining criteria*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC. Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/presentation-2012-3-developing-performance-level-descriptors-criteria.pdf

Camara, W. J. (2013). Defining and measuring college and career readiness: A validation framework. *Educational Measurement: Issues and Practice, 32*(4), 16–27.

Camara, W., & Quenemoen, R. (2012). *Defining and measuring college and career readiness and informing the development of performance level descriptors*. Retrieved from http://www.parcconline.org/sites/parcc/files/PARCC%20CCR%20paper%20v14%201-8-12.pdf

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31–44.

Conley, D. T., Drummond, K. V., Gonzalez, A., Rooseboom, J., & Stout. O. (2011). *Reaching the goal: The applicability and importance of the Common Core State Standards to college and career readiness*. Eugene, OR: Educational Policy Improvement Center.

Crocker, L. M., Miller, D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. *Applied Measurement in Education*, *2*, 179–194.

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum.

Crotts, K., Sireci, S. G., & Zenisky, A. L. (2012). Evaluating the content quality of a multistage-adaptive test. *Journal of Applied Testing Technology, 13*(1), 1–26.

D'Agostino, J. V., & Bonner, S. M. (2009). High school exit exam scores and university performance. *Educational Assessment, 14*, 25–47.

D'Agostino, J., Karpinski, A., & Welsh, M. (2011). A method to examine content domain structures. *International Journal of Testing*, *11*(4), 295–307.

D'Agostino, J. V., Welsh, M., & Corson, N. M. (2007). Instructional sensitivity of a state's standards-based assessment. *Educational Assessment, 12*, 1–22.

Davis-Becker, S. L., Buckendahl, C. W., & Gerrow, J. (2011). Evaluating the bookmark standard setting method: The impact of random item ordering. *International Journal of Testing*, *11*(1), 24–37.

Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial and ethnic minorities. *American Psychologist*, 53, 728–736.

Delisle, D. S. (2012). *Letter to chief state school officers*. Washington, DC: U.S. Department of Education.

Dorans, N. J. (2004). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement, 41*, 43–68.

Duncan, G. D., del Rio Parant, L., Chen, W.-H., Ferrara, S., Johnson, E., Oppler, S., & Shieh, Y.-Y. (2005). Study of a dual-language test booklet in eighth-grade mathematics. *Applied Measurement in Education, 18*, 129–161.

Educational Testing Service (ETS). (2012a). *Smarter09 component 1: Narrative key to understanding the blueprint tables (TAC draft)*. Princeton, NJ: Author.

Educational Testing Service (ETS). (2012b). *Smarter Balanced Assessment Consortium: Bias and sensitivity guidelines*. Princeton, NJ: Author.

Educational Testing Service (ETS). (2012c). *Specifications for an interim system of assessment*. Princeton, NJ: Author.

Embretson (Whitley), S. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179–197.

Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing, 5*(1), 23–35.

Fields, R., & Parsad, B. (2012). *Tests and cut scores used for student placement in postsecondary education: Fall 2011*. Washington, DC: National Assessment Governing Board.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data. *School Psychology Review, 29,* 65–85.

Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice, 18*(4), 5–9.

Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. A. Berk (Ed.), *Criterion-referenced measurement: The state of the art*. Baltimore, MD: Johns Hopkins University Press.

Hambleton, R. K. (1989). Principles and applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 147–200). New York, NY: Macmillan.

Hambleton, R. K. (2005). Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting psychological and educational tests for cross-cultural assessment* (pp. 3–38). Hillsdale, NJ: Lawrence Erlbaum.

Hambleton, R. K., & Han, N. (2004). *Summary of our efforts to compute decision consistency and decision accuracy estimates using IRT item statistics* (Center for Educational Assessment

Research Report No. 552). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Hambleton, R. K., & Rovenelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement*, *10*, 287–302.

Hambleton, R. K., & Slater, S. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). Los Angeles, CA: National Center for Research on Evaluation, Standards, & Student Testing.

Hamilton, L. S. (1994, April). *Validating hands-on science assessments through an investigation of response processes*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of a set of test items. *Applied Psychological Measurement*, 9, 139–164.

Hill, R. K., & DePascale, C. A. (2002). *Determining the reliability of school scores*. Dover, NH: National Center for the Improvement of Educational Assessment.

Hill, R. K., & DePascale, C. A. (2003, April). *Adequate yearly progress under NCLB: Reliability considerations*. Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, IL.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

International Test Commission (ITC). (2010). *Guidelines for translating and adapting tests*. Retrieved from http://www.intestcom.org

Johnstone, C. J., Altman, J. M., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/nceo/OnlinePubs/StateGuideUD/default.htm

Kaira, L. T., & Sireci, S. G. (2010). Evaluating content validity in multistage adaptive testing. *CLEAR Exam Review*, *21*(2), 15–23.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527–535.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64,* 425–461.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 53–88). Mahwah, NJ: Erlbaum.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: American Council on Education/Praeger.

Keng, L., Murphy, D., & Gaertner, M. (2012, April). *Supported by data: A comprehensive approach for building empirical evidence for standard setting*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Kolen, M. J. (2011). *Issues associated with vertical scales for PARCC assessments*. Retrieved from http://www.parcconline.org/sites/parcc/files/PARCCVertScal289-12-201129.pdf

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

La Salle, A., Munoz, C., Ruff, L., Weisman, E., Sedillo, R., & Phillips, L. (2012, April). *Grounded in the content: The role of content analysis in evidence-based standard setting*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Lee, W. (2008). *Classification consistency and accuracy for complex assessments using item response theory* (CASMA Research Report No. 27). Iowa City, IA: University of Iowa.

Leighton, J. P. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice, 23*(4), 6–15.

Liang, T., Han, K. T, & Hambleton, R. K. (2008). *User's guide for ResidPlots-2: Computer software for IRT graphical residual analyses, Version 2.0* (Center for Educational Assessment Research Report No. 688). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Liang, T., Han, K. T., & Hambleton, R.K. (2009). ResidPlots-2: Computer software for IRT graphical residual analyses. *Applied Psychological Measurement, 33*(5), 411–412.

Linacre, J. M. (2004). *A user's guide to Winsteps Rasch-model computer programs*. Chicago, IL: MESA Press.

Linn, R. L. (2006). The standards for educational and psychological testing: Guidance in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 27–38). Mahwah, NJ: Lawrence Erlbaum.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3–16.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197.

Loomis, S. C. (2011, April). *Toward a validity framework for reporting preparedness of 12th graders for college-level course placement and entry to job training programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., Sheinker, J., & Council of Chief State School Officers (CCSSO). (2002). *Making valid and reliable decisions in determining adequate yearly progress. A paper in the series: Implementing the State Accountability System Requirements under the No Child Left Behind Act of 2001*. Washington, DC: CCSSO.

Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the physical self description questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology*, *32*(4), 438–482.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction. *Review of Educational Research, 4*, 1332–1361.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education.

Mislevy, R. J. (2009). *Validity from the perspective of model-based reasoning* (CRESST Report 752). Los Angeles, CA: National Center for Research on Evaluation, Standards, & Student Testing.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 61–90), Mahwah, NJ: Lawrence Erlbaum.

National Assessment Governing Board (NAGB). (2010). *Program of preparedness research study brief.* Retrieved from http://www.nagb.org/content/nagb/assets/documents/newsroom/press-releases/2010/release-20101122/research.pdf

National Council on Measurement in Education (NCME). (2012). *Testing and data integrity in the administration of statewide student assessment programs*. Madison, WI: Author.

National Governors Association Center for Best Practices (NGA Center) & CCSSO. (2010). *Common Core State Standards for mathematics*. Washington, DC: Authors.

Northwest Evaluation Association. (2005). *Technical manual: For use with Measures of Academic Progress and achievement level tests*. Lake Oswego, OR: Author.

O'Malley, K., Keng, L., & Miles, J. (2012). Using validity evidence to set performance standards. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed.) (pp. 301–322). New York, NY: Routledge.

O'Neil, T., Sireci, S. G., & Huff, K. F. (2004). Evaluating the consistency of test content across two successive administrations of a state-mandated science assessment. *Educational Assessment, 9*, 129–151.

Patz, R. J. (2007). *Vertical scaling in standards-based educational assessment and accountability systems*. Washington, DC: CCSSO.

Penfield, R. D., & Miller, J. M. (2004). Improving content validation studies using an asymmetric confidence interval for the mean of expert ratings. *Applied Measurement In Education*, *17*(4), 359-–370.

Pitoniak, M. J., Sireci, S. G., & Luecht, R. M. (2002). A multitrait-multimethod validity investigation of scores from a professional licensure exam. *Educational and Psychological Measurement, 62*, 498–516.

Popham, W. J. (1992). Appropriate expectations for content judgments regarding teacher licensure tests. *Applied Measurement in Education, 5*, 285–301.

Porter, A. C., & Smithson, J. L. (2002, April). *Alignment of assessments, standards and instruction using curriculum indicator data*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Rabinowitz, S., Zimmerman, J., & Sherman, K. (2001). *Do high-stakes tests drive up student dropout rates? Myths versus realities* (Research brief). San Francisco, CA: WestEd.

Ramsey, P. A. (1993). Sensitivity review: The ETS experience as a case study. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 367–388). Hillsdale, NJ: Erlbaum.

Reckase, M. D. (2006a). A conceptual framework for a psychometric theory for standard setting with examples of its use for evaluating the functioning of the standard setting methods. *Educational Measurement: Issues and Practice*, *25*(2), 4–18.

Reckase, M. D. (2006b). Rejoinder: Evaluating standard setting methods using error models proposed by Schultz. *Educational Measurement: Issues and Practice*, *25*(3), 14–17.

Robin, F., Sireci, S. G., & Hambleton, R. K. (2003). Evaluating the equivalence of different language versions of a credentialing exam. *International Journal of Testing, 3*, 1–20.

Rothman, R. (2003). *Imperfect matches: The alignment of standards and tests*. Washington, DC: National Research Council.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment, Research & Evaluation*, *7*(14).

Rudner, L. M. (2004). *Expected classification accuracy*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Select Committee. (1994/1995). Sampling and statistical procedures used in the California Learning Assessment System. In L. J. Cronbach (Ed.). *A valedictory: Reflections on 60 years in educational testing* (Board Bulletin). Washington, DC: National Research Council, Board on Testing and Assessment.

Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405–450.

Silk, Y., Silver, D., Amerian, S., Nishimura, C., & Boscardin, C. K. (2009). *Using classroom artifacts to measure the efficacy of a professional development* (CRESST Report 761). Los Angeles, CA: National Center for Research on Evaluation, Standards, & Student Testing.

Simpson, M., Gong, B., Marion, S., National Center on Educational Outcomes, CCSSO, & National Association of State Directors of Special Education. (2006). *Effect of minimum cell sizes and confidence interval sizes for special education subgroups on school-level AYP determinations* (Synthesis Report 61). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Sireci, S. G. (1997). Problems and issues in linking tests across languages. *Educational Measurement: Issues and Practice, 16*(1), 12–19.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*, 299–321.

Sireci, S. G. (2009). Packing and upacking sources of validity evidence: History repeats itself again. In R. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19–37). Charlotte, NC: Information Age Publishing Inc.

Sireci, S. G., & Geisinger, K. F. (1992). Analyzing test content using cluster analysis and multidimensional scaling. *Applied Psychological Measurement, 16,* 17–31.

Sireci, S. G., & Geisinger, K. F. (1995). Using subject matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement, 19*, 241–255.

Sireci, S. G., Han, K. T., & Wells, C. S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment, 13*, 108–131.

Sireci, S. G., Hauger, J. B., Wells, C. S., Shea, C., & Zenisky, A. L. (2009). Evaluation of the standard setting on the 2005 grade 12 National Assessment of Educational Progress mathematics test. *Applied Measurement in Education, 22*, 339–358.

Sireci, S. G., & Mullane, L. A. (1994). Evaluating test fairness in licensure testing: The sensitivity review process. *CLEAR Exam Review, 5*(2), 22–28.

Sireci, S. G., Robin, F., Meara, K., Rogers, H. J., & Swaminathan, H. (2000). An external evaluation of the 1996 grade 8 NAEP science framework. In N. Raju, J. W. Pellegrino, M. W. Bertenthal, K. J. Mitchell, & L. R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP* (pp. 74–100). Washington, DC: National Academies Press.

Sireci, S. G., Rogers, H. J., Swaminathan, H., Meara, K., & Robin, F. (2000). Appraising the dimensionality of the 1996 grade 8 NAEP science assessment data. In N. Raju, J. W. Pellegrino, M. W. Bertenthal, K. J. Mitchell, & L. R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP* (pp. 101–122). Washington, DC: National Academies Press.

Sireci, S. G., & Schweid, J. A. (2011, April). *Beyond alignment: Important questions to ask (and answer) to evaluate content validity*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Sireci, S. G., & Wells. C. S. (2010). Evaluating the comparability of English and Spanish video accommodations for English language learners. In P. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 33–68). Washington, DC: CCSSO.

Smarter Balanced Assessment Consortium (Smarter Balanced). (2010). *Race to the Top assessment program application for new grants: Comprehensive assessment systems, CFDA Number: 84.395B*. OMB Control Number 1810-0699.

Smarter Balanced Assessment Consortium (Smarter Balanced). (2012a). *Master work plan—formative*. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/03/Formative-Assessment-Master-Work-Plan-Narrative.pdf

Smarter Balanced Assessment Consortium (Smarter Balanced). (2012b). *Theory of action. An excerpt from the Smarter Balanced Race to the Top application*. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/02/Smarter-Balanced-Theory-of-Action.pdf

Smarter Balanced Assessment Consortium (Smarter Balanced). (n.d.). *Frequently asked questions*. Retrieved from http://www.smarterbalanced.org/resources-events/faqs/

Tomlinson, M. R., & Fortenberry, N. (2008, October). *Classroom artifacts: Tools to assess the use of active, innovative, and engineering pedagogies among engineering faculty*. Paper presented at the annual ASEE/IEEE Frontiers in Education conference, Saratoga Springs, NY. Retrieved from http://fie-conference.org/fie2008/papers/1088.pdf

Tong, Y., & Kolen, M. J. (2007). Comparisons of methodologies and results in vertical scaling for educational achievement tests. *Applied Measurement In Education*, *20*(2), 227–253.

U.S. Department of Education. (2009a). *Race to the Top program executive summary*. Washington, DC: Author.

U.S. Department of Education. (2009b). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001. [Revised December 21, 2007 to include modified academic achievement standards. Revised with technical edits, January 12, 2009]*. Washington, DC: Author.

U.S. Department of Education (2010). *U.S. secretary of education Duncan announces winners of competition to improve student assessments*. Retrieved from http://www.ed.gov/news/press-releases/us-secretary-education-duncan-announces-winners-competition-improve-student-asse

Vasavada, N., Carman, E., Hart, B., & Luisser, D. (2010). *Common Core State Standards alignment: ReadiStep, PSAT/NMSQT, and SAT* (Research Report 2010-5). New York, NY: The College Board.

Wainer, H. (2011). *Uneducated guesses: Using evidence to uncover misguided education policies.* Princeton, NJ: Princeton University Press.

Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement, 36*, 301–335.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Washington, DC: CCSSO.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7–25.

Welch, C., & Dunbar, S. (2011, April). *K–12 assessments and college readiness: Necessary validity evidence for educators, teachers, and parents*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Wells, C. S., Baldwin, S., Hambleton, R. K., Sireci, S. G., Karantonis, A., & Jirka, S. (2009). Evaluating score equity assessment for state NAEP. *Applied Measurement in Education, 22,* 394–408.

Wilson, D, Wood, R., & Gibbons, R. D. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis* [computer program]. Mooresville, IN: Scientific Software.

Williams, N. J., Keng, L., & O'Malley, K. (2012, April). *Maximizing panel input: Incorporating empirical evidence in a way the standard setting panel will understand*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *SAT benchmarks: Development of a college readiness benchmark and its relationship to secondary and postsecondary school performance*. (Research Report 2011-5). New York, NY: The College Board.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Yen, W. M. (1997). The technical quality of performance assessments: Standard errors of percents of pupils reaching standards. *Educational Measurement: Issues and Practice, 16*(3), 5–15.

Young, J., Pitoniak, M. J., King, T. C., & Ayad, E. (2012). *Smarter Balanced Assessment Consortium: Guidelines for accessibility for English language learners*. Retrieved from http://www.smarterbalanced.org/smarter-balanced-assessments/

Zwick, R., & Schlemer, L. (2004). SAT validity for linguistic minorities at the University of California, Santa Barbara. *Educational Measurement: Issues and Practice*, *23*(1), 6–16.

## Appendix A: Smarter Balanced Theory of Action and Derivation of Purpose Statements

### Smarter Balanced Assessment Consortium Theory of Action

Bennett (2010) described a Theory of Action (TOA) as follows:

> *Theory of Action* is a common notion in the program evaluation literature . . . appearing to have come about because program managers were too often unclear about the intended goals of their efforts. The term is closely associated with *logic model*, a graphical or textual description of an intervention that explains the cause-effect relationships among inputs, activities, and intended outcomes. (pp. 70-71)

Smarter Balanced's TOA is well articulated in its Race to the Top application (Smarter Balanced, 2010) and has been excerpted from the application as a separate document available on the SBAC website (Smarter Balanced, 2012b). It begins by stating that Smarter Balanced "supports the development and implementation of learning and assessment systems to radically reshape the education enterprise . . . to improve student outcomes" and states that "the overarching goal of the Smarter Balanced Assessment Consortium is to *ensure that all students leave high school prepared for postsecondary success in college or a career through increased student learning and improved teaching*" (p. 1; emphasis in original). The TOA lists "seven principles undergirding the theory of action" (p. 1). These principles are:

1. Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of an integrated system of standards, curriculum, assessment, instruction, and teacher development.

2. Assessments produce evidence of student performance on challenging tasks that evaluate the Common Core State Standards.

3. Teachers are integrally involved in the development and scoring of assessments.

4. The development and implementation of the assessment system is a state-led effort with a transparent and inclusive governance structure.

5. Assessments are structured to continuously improve teaching and learning.

6. Assessment, reporting, and accountability systems provide useful information on multiple measures that is educative for all stakeholders.

7. Design and implementation strategies adhere to established professional standards. (Smarter Balanced, 2010, pp. 32–33)

From these principles we can immediately infer that intended goals of Smarter Balanced are to develop quality assessments that are aligned with the CCSS, are part of a system that supports instruction and student learning, and provide results that are useful for evaluating student performance. It is also clear that other goals are to involve teachers throughout the test development and scoring processes and to operate as a true collaborative with states working in unison toward these common goals.

The model that Smarter Balanced established to meet these goals involves three different components: (a) summative assessments, (b) interim-benchmark assessments, and (c) formative assessment resources. A schematic representation of the Smarter Balanced TOA is illustrated in Figure A-1, which is taken directly from the Smarter Balanced Race to the Top application (Smarter Balanced, 2010). This representation includes the three assessment components, but also illustrates the other components that are required for the Consortium members to work together in unison and to reach the "overarching goal" found on the right side of the figure. Related to the Theory of Action are the overall and specific claims for the summative assessments, which are presented in Table A-1.

Figure A-1. Overview of Smarter Balanced Assessment Consortium Theory of Action
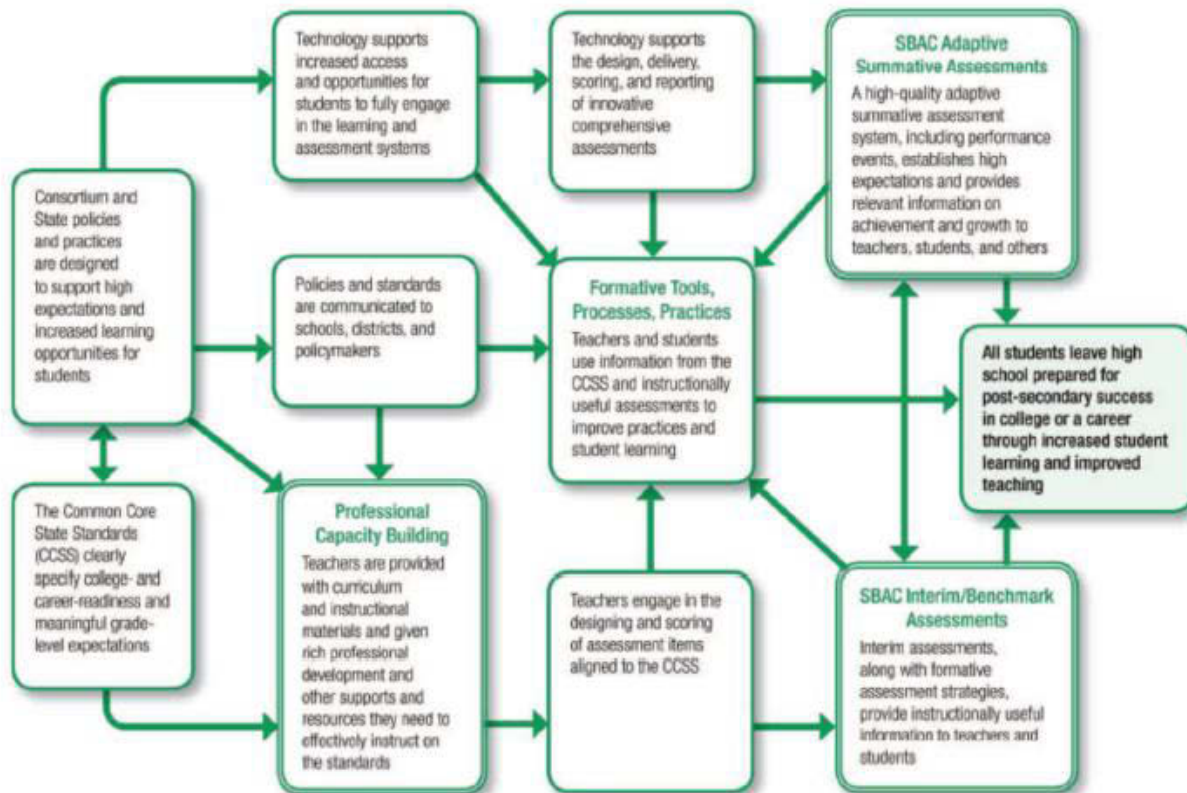


*Source: Smarter Balanced (2012b).*

Table A-1. Overall and Specific Claims for Smarter Balanced Summative Assessments

| Claim Type | ELA: Students can . . . | Mathematics: Students can . . . |
|---|---|---|
| Overall: Grades 3–8 | demonstrate progress toward college and career readiness in English language arts and literacy. | demonstrate progress toward college and career readiness in mathematics. |
| Overall: Grade 11 | demonstrate college and career readiness in English language arts and literacy. | demonstrate college and career readiness in mathematics. |
| Specific | read closely and analytically to comprehend a range of increasingly complex literary and informational texts. | explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency. |
| | produce effective and well-grounded writing for a range of purposes and audiences. | solve a range of complex, well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. |
| | employ effective speaking and listening skills for a range of purposes and audiences. | clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. |
| | engage in research and inquiry to investigate topics, and to analyze, integrate, and present information. | analyze complex, real-world scenarios and construct and use mathematical models to interpret and solve problems. |

## Appendix B: Description of Alignment Methods

| Alignment Model | Dimension | Brief Description |
|---|---|---|
| Webb (1997) | *Categorical Concurrence | Match of items to general content areas |
| | **Depth of Knowledge Consistency | Cognitive level of items compared to cognitive level of benchmark/objective |
| | **Range of Knowledge Correspondence | Number of benchmarks/objectives measured within general content area |
| | **Balance of Representation | Distribution of items across general content areas |
| Achieve (2006) | *Content Centrality | Congruence between item and objective/benchmark |
| | *Performance Centrality | Congruence between cognitive demand of item and objective/benchmark |
| | **Source of Challenge | Grade-level appropriateness |
| | *Level of Cognitive Demand | Cognitive level measured by item |
| | **Level of Challenge | Degree to which test captures difficulty implied by general content areas |
| | **Balance | Holistic evaluation of how well test represents content/cognitive specs |
| | **Range | Proportion of objectives/benchmarks measured within general content area |
| SEC (Porter et al., 2001) | *Content Match | Match of items to content areas and cognitive levels |
| | **Expectations for Student Performance | Compares cognitive demands of curriculum and assessment |
| | **Instructional Content | Compares what is taught with what is tested |

*Covered or partially covered by one or more traditional content validation approaches.
**Unique contribution of alignment method.
From Sireci & Schweid (2011).

## Appendix C: Description of Item Similarity Rating Approach to Evaluating Test Content

As stated earlier, a disadvantage of this approach to blueprint confirmation is that it may foster social desirability—that is, by informing SMEs of the intended CCSS measured by each item, it may unconsciously bias their ratings in support of item/standard congruence. To avoid this potential confound, and to determine whether other relations among the items are present that are not described in the test specifications, the item similarity rating task described earlier could be conducted. An example of this task is presented in Figure C-1. An example of some of the results from this type of study (from Sireci, Robin, Meara, Rogers, & Swaminathan, 2000) is presented in Figure C-2. These results could be followed up by cluster analyses, to see if the items cluster as intended by the test specifications.
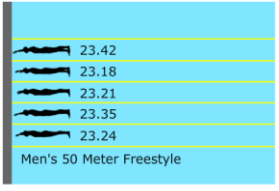
Given that the item similarity rating task requires more SME time and more complex data analysis, we recommend that all items be rated for congruence using an alignment-type rating task similar to that illustrated in Exhibit 1. However, the similarity rating procedure provides a more stringent test and protects against confirmationist bias (social desirability), and so should be considered as a supplementary study, perhaps using a subset of items.

Figure C-1. Example of Item Similarity Rating Task

Directions: Please review each pair of items and rate how similar the two items are to one another in terms of the mathematics knowledge and skills measured using the rating scale provided.
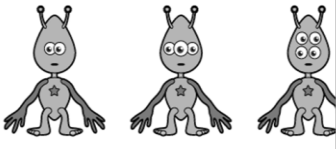


| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

Very Similar                                                                                     Very Different

Figure C-2. Example of Results from Item Similarity Ratings Study
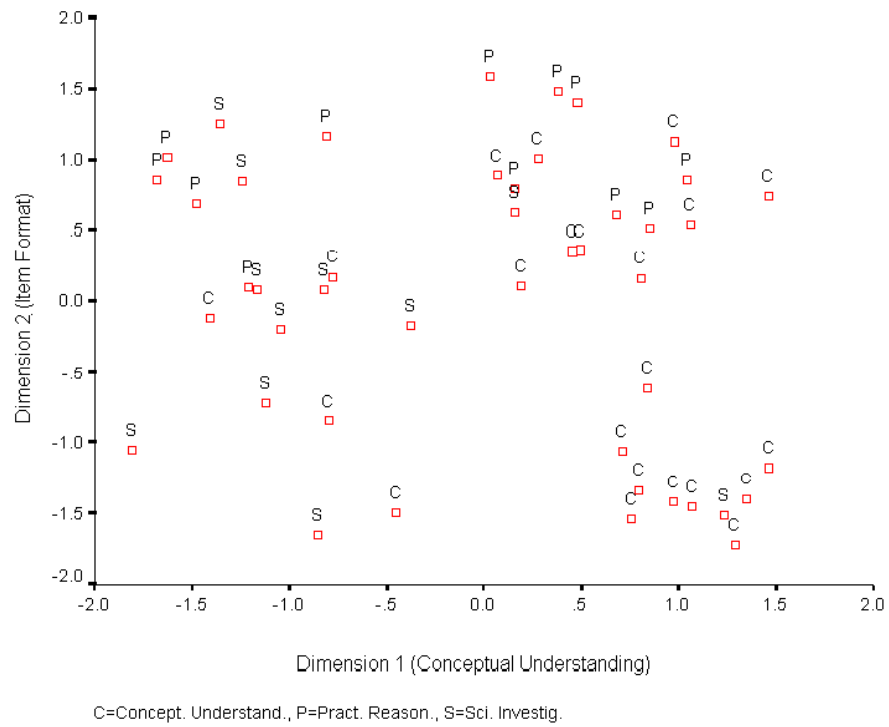


Source: Sireci et al., 2000.

# Appendix D: Description of ResidPlots2: IRT Residual Analysis Software

ResidPlots-2 (Liang, Han, & Hambleton, 2008, 2009) is a software program for evaluating the fit of item response theory (IRT) models to data. By comparing observations to model-predicted expectations, ResidPlots-2 works at the item level to provide researchers with information to determine how well an IRT model fits a given data set. The approach used in ResidPlots-2 is to first compute model fit statistics using the observed data, and then also use item and ability estimates from IRT estimation programs, such as BILOG-MG, PARSCALE, and MULTILOG, to simulate examinee response data and report the average from 10 replications of the simulation. Thus, simulation results obtained in this way better approximate the expected observed test score distribution.

The output from ResidPlots-2 takes the forms of both graphs and tables. Plots generated by ResidPlots-2 include:

- Item-level plots (raw residual plots, standardized residual plots),
- Test-level plots (standardized residual distributions [both cumulative density function {CDF} and probability density function {PDF}], item and score fit plots from empirical and simulated data); and
- Score plots (observed and predicted test score distributions).

ResidPlots-2 also generates six tables of results:

- The *FIT STAT* table provides results for two fit statistics at the item level (chi square, G square) as well as degree of freedom and fit probability for both, and basic item details (item number, parameter estimates, and sample size).
- The *SR PDF* table lists details of the standardized residual (SR) distribution for the PDF, with mean, standard deviation, and relative frequency of the SR distribution. These results are provided for the overall test and broken out by format (dichotomous and polytomous items) and for both observed and simulated data.
- The *SR CDF* table is a companion table to the SR PDF table; here, the results are provided for the CDF.
- The *NCOUNT* table displays the characteristics of the sample (sample size and percentage) in each reported interval for each item. This is an important feature, as users can make application-specific choices about interval width and score ranges in ResidPlots-2.
- The *PFIT* table provides the results of the Lz person fit statistic for each person in the sample. Note that this report lists the probability values for each person, where values below 0.05 are indicative of person misfit.
- The *P_RISE* table contains results for the root integrated square error statistic (RISE), which is a nonparametric fit statistic. As with the PFIT table, results are shown in terms of probability values for each item, where values less than 0.05 are indicative of nonparametric item misfit.

The plots in Figures D-1 and D-2 are samples of output from ResidPlots-2 that depict the item-fit plot. Note that the 3P model was fit to the data for Figure D-1, while a 1P model was fit to the same data for Figure D-2. Figure D-2 illustrates that results from the observed calibration are much more disparate from the simulated results than the results shown in Figure D-1, which suggests that the 3P model provides better model-data fit than the 1P model for the data.

Figure D-1. ResidPlots-2 Item Fit Plot (data fit by 3P)
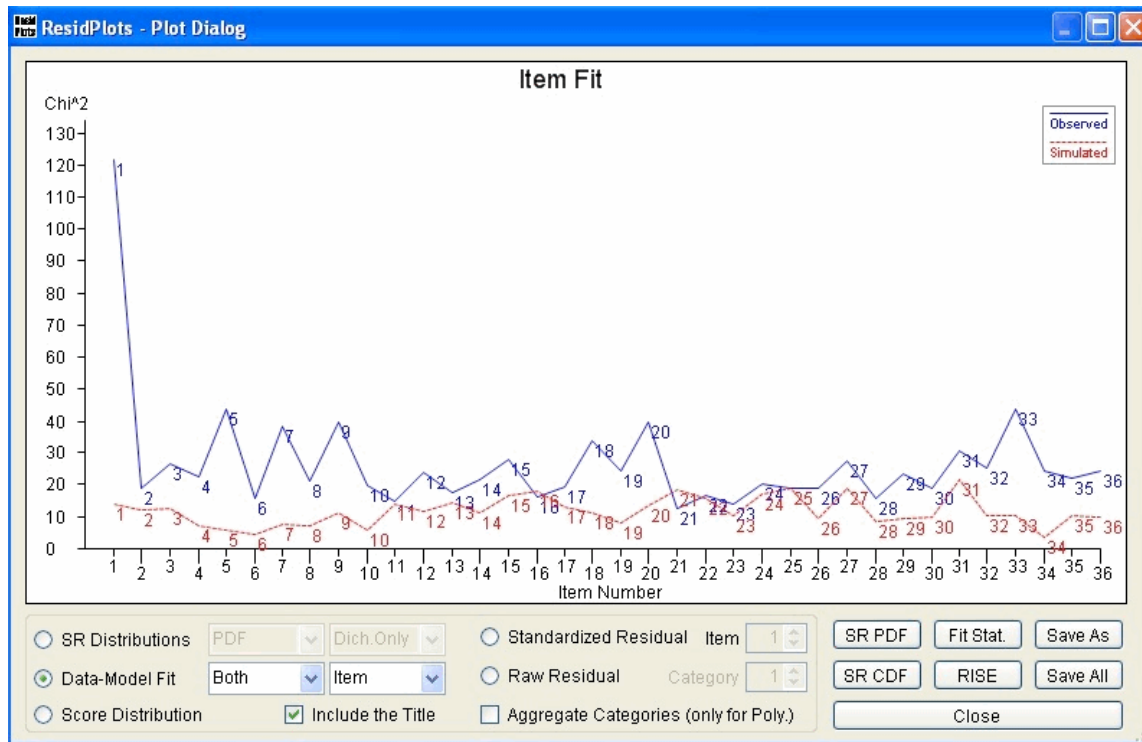


Figure D-2. ResidPlots-2 Item Fit Plot (data fit by 1P)

Appendix F –     Cognitive Lab Final Report
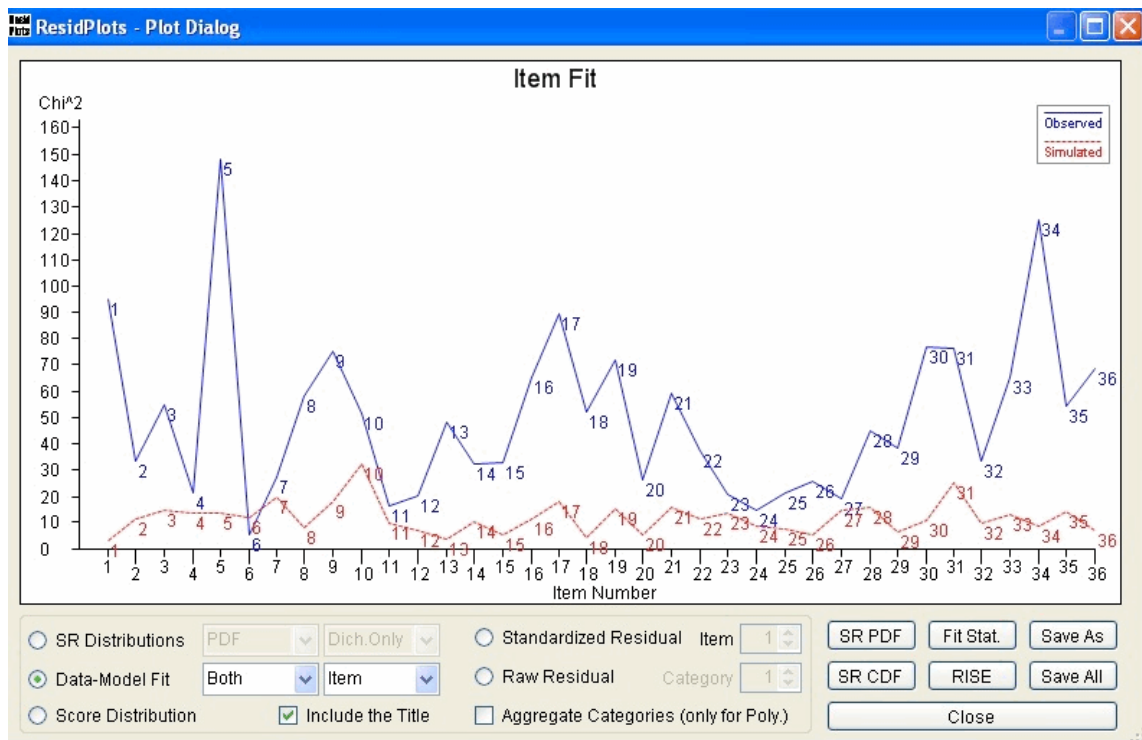
# Smarter Balanced Assessment Consortium:

## Cognitive Laboratories Technical Report

Developed by: The American Institutes for Research

September 27, 2013

Executive Summary

The Smarter Balanced Assessment Consortium conducted cognitive laboratories to better understand how students solve various types of items. A cognitive laboratory uses a think-aloud methodology in which students speak their thoughts while solving a test item. The interviewer follows a standardized protocol to elicit responses and record what a student says. While this one-on-one process is time consuming, the type of information elicited is often difficult to obtain by other means. This report presents the results of a series of cognitive laboratory observational studies. The studies were conducted with small numbers of students in order to gather in-depth qualitative data about how students react to different types of items, formats, etc. Due to the small number of subjects studied and the ad hoc nature of the achieved sample of participants, the findings should be used to point the way to more systematic studies, rather than be cited as an authoritative source of scientific findings.

This executive summary presents the major findings from various protocols. Most protocols were developed at multiple grade bands (e.g., 3, 6, and 11). A grade band is the level of content for which the protocol is targeted. Protocols were usually targeted to answer a specific question in one or more content areas (e.g., ELA, mathematics). Results are organized under topics or questions of interest.

## Summary and Findings of Cognitive Lab Results by Research Question

*Research Question 1: Do mathematics multi-part selected-response (MPSR) items provide similar information about the depth of understanding by the test taker similar to traditional constructed-response (CR) items?*

An MPSR item has students select several examples of a correct response rather than just one, as in the typical selected-response (SR) item. The intention of this research question was to see whether the MPSR items provided depth of understanding similar to that provided by CR items. If effective, an MPSR item would be a more efficient way to measure the content measured by CR items. Within a form, parallel items were constructed in both formats and presented to the same students. In the protocols the MPSR and CR items were presented in random order.

This research question sought to address two hypotheses. The first hypothesis examined whether students who get full credit on MPSR items reveal, through their think-aloud sessions, greater understanding than those students who do not achieve full credit. The second hypothesis examined whether students who get full credit on MPSR items reveal depth of understanding similar to that of students who get full credit on similarly challenging CR items measuring the same target.

In most cases, the depth of knowledge (DOK) demonstrated by the student for the MPSR items either equaled or exceeded the DOK demonstrated for the CR items. Students who got full credit on the MPSR items also revealed greater understanding of the material than those who did not obtain full credit. The percentage of students understanding the material was also quite similar for the

MPSR and CR items. A typical interviewer comment was, "based on the accuracy of the student's responses to both types of items, it appears that item type is not a factor in determining how well the students respond[s]."

*Research Question 2: Under what conditions do specific types of TE items (and SR items) approach the depth of knowledge (DOK) of a written constructed response in ELA and mathematics?*

This question was designed to assess whether different types of technology-enhanced (TE) items approach the DOK of CR items for specific content claim/targets and DOK levels. SR items were also included, where available, as a comparison item format. Comparisons were examined for specific TE item types at specific DOK levels for specific content claims/targets. CR and SR items were matched to specific content claims/targets and DOK 4 items in one of the three formats (SR, TE, and CR) appeared in each form. Multiple forms were administered, each form to a different sample of students. It was hypothesized that students responding to items of a specific type would reveal that they were using thought processes consistent with a specific DOK level for items measuring a specific target. Different item types were administered to different students.

For ELA, a higher percentage of students demonstrated thought processes consistent with the specific DOK levels for most of the TE item types than for the attached CR items. Two exceptions were two targets in the "select text" item type: "justifying interpretations" (grade band 6) and "analyzing the figurative" (grade band 11). A similar pattern was observed for the matched SR items versus the CR items.

Regarding student performance on the ELA items, the pattern of results were very similar to that observed for the DOK consistency-of-thought processes.  The same TE item types had higher percentages of students receiving the maximum score than did the matched CR items with the exception of the "select text" items for the "writing or revising strategies" target (grade band 7) and the "citing to support inferences" target (grade band 11).

For the SR items in ELA, the percentage receiving the maximum score was higher than both the CR and TE formats for the following "select text" items:

- "select text" for justifying interpretations, claim 1, DOK 2 in grade band 6
- "select text" for citing to support inferences, claim 1, DOK 2 in grade band 11
- "select text" for analyzing the figurative, claim 1, DOK 2 in grade band 11

For mathematics, the results were more varied.  Compared to the matched CR items, the following TE item types had a higher percentage of students demonstrating thought processes consistent with the DOK level.

- "placing points" for fractions, claim 1, DOK 2 in grade band 3
- "single lines" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "tiling" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11 ("Student indicated use of multiple steps and solved correctly.")
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 4, DOK 3 in grade band 4

The item types in which the CR items had a higher percentage of DOK-consistent thought processes included:

- "select and order" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6
- "tiling" for everyday mathematic problems, claim 4, DOK 3 in grade band 4
- "tiling" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6
- "tiling" for everyday mathematic problems, claim 2, DOK 3 in grade band 11
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4

The TE item types for which a higher percentage of students received full credit included only:

- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11, and
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4.

In other cases, the percentage of students receiving full credit was lower for the TE item types than for the comparable CR items. It should be noted that the percentage receiving full credit was generally low in mathematics for all three item formats. Even the matched SR items generally did not perform any better than either the CR or TE items.

*Research Question 3: For multi-part selected response (MPSR) items where students may select more than one answer choice, which wording best indicates to the student that he or she is allowed to select more than one option? For multipart dichotomous-choice (e.g., YES/NO) items, do students know that they need to answer each part?*

Smarter Balanced sought to investigate whether students might become confused with MPSR items in mathematics and perhaps not complete the entire item. In order to investigate this, items were constructed with different amounts of labeling. *Labeling* is the identification of the parts of the problem with indicators such as "a," "b," "c" or "1," "2," "3." For each MPSR item, labeled and non-labeled condition were investigated. An example of an item in labeled and non-labeled format can be found in Exhibit 1.

This question was designed to assess whether labeling or not labeling an MPSR mathematics item produces a difference in performance. Forms were constructed for five grade bands, with each form containing one MPSR item followed by one CR item. The labeled and non-labeled items appeared in different forms of the test and thus were taken by different students.

Even though the labeling of MPSR items was intended to clarify the mathematic tasks for the students, in many cases it actually seemed to confuse the students. Little difference was observed between the labeled and non-labeled items in the lower grade bands (grade bands 3–6). However, students in grade band 7 tended to score higher with non-labeled items. Also, students in grade bands 7 and 11 tended to be confused by the labeling. In addition, the labeled items tended to receive more comments related to not understanding the instructions. The interviewer confirmed this, suggesting that the grade bands 7 and 11 students better understood the instructions in the non-labeled condition than in the labeled condition.

*Research Question 4: Does the ability to move one or more sentences to different positions provide evidence of students' ability to revise text appropriately in the consideration of chronology, coherence, transitions, or the author's craft?*

Smarter Balanced is considering using ELA items that have students reorder sentences to measure an editing/revising standard. Claim 2 of the ELA standards states that students should be able to revise one or more paragraphs demonstrating specific narrative strategies (use of dialogue, sensory or concrete details, description), chronology, appropriate transitional strategies for coherence, or authors' craft appropriate to the purpose of the item (closure, detailing characters, plot, setting, or an event).

This question was designed to assess whether students' movement of one or more sentences to different positions provided evidence of students' ability to demonstrate consideration of chronology, coherence, transitions, or author's craft. Six ELA items were included in a test form.

Students who performed well on the items were more likely to consider the targeted writing skills (e.g., chronology, coherence, transitions, and author's craft) when answering the questions. The results showed that students who made more appropriate sentence moves (and fewer inappropriate moves) were more likely to consider the writing skills of chronology, coherence, and transitions. The pattern was less clear for consideration of author's craft.

*Research Question 5: Do students who construct text reveal more understanding of targeted writing skills than students who manipulate writing through the manipulation of text (MT) tasks?*

Many believe that the best way to measure writing is to have students write. However, in a testing environment, it is often difficult to adequately sample the writing content domain with an assessment composed exclusively of CR items. An effort is ongoing to find items that are efficient but that can adequately measure the components of the writing domain, thus allowing a broader selection and greater number of items to be delivered. This question examined whether students responding to MT tasks would demonstrate understanding of the targeted writing skills comparable to the understanding demonstrated for CR tasks assessing the same claim and target. Examples of the item types can be found in Exhibit 2.

For each of three grade bands (3, 6, and 11), four pairs of ELA items were developed. Two forms were created for each grade band, and each pair contained one MT item and one CR version of the same item. Two forms were created, and each form contained a single version of an item. Each form contained two MT items and two CR items. The MT items were almost exclusively "select and order" items, though two items—one in grade band 3 and one in grade band 11—were "reorder text" items. All items assessed claim 1, target 1.

The results showed that the targeted writing skills are considered by students who manipulate text at a level comparable to (or greater than) that encountered when they are constructing text. The students in grade bands 3 and 6 showed comparable (or greater) levels of understanding when the items were in an MT format. For the grade band 11 students, the results were mixed, but students tended to be more effective in applying the targeted writing skills in the CR format, particularly for transitions and author's craft. Score distributions were comparable for MT and CR item formats.

*Research Question 6: Do different types of directions (minimal, concise, or extensive) have an effect on the performance of technology enhanced (TE) items in ELA and Mathematics?*

The optimal amount of direction that should be given to a student working with TE items is unclear. With minimal directions, students may not know how to approach an item; with extensive directions, students may be distracted or slowed to a point where the item becomes inefficient. This may be particularly true with elementary school students, who may take longer to process text. This question examined this issue for ELA and mathematics items. Three types of directions were used (minimal, concise, and extensive).

In most cases in ELA, the level of instruction did not make a difference. For most grade bands and item types, neither the level of instruction nor the item type showed a differential effect in ELA. Cases in which differences were observed included "select text" items when the directions were "concise." With the "reorder text" items, the grade band 3 students did less well with minimal directions. The grade band 11 students also had some difficulty with the "reorder text" items when the directions were "extensive."

In mathematics, the level of instruction also did not make a difference for many item types and grade bands. "Select and order" items were difficult (grade bands 6 and 11) regardless of the direction type;however, no direction type proved better than another. High percentages of students received full credit on "select defined partition" and "straight lines" items; however, the direction type did not make a difference. Finally, "tiling" items were generally difficult, but no benefit was shown for different types of directions. Differences were observed in items including "placing points" items under the minimal and concise directions in grade band 11; however, under extensive directions, all students received the maximum score. With "placing points and tiling" items, a higher percentage of students received full credit with fewer instructions (grade band 6). Finally, "vertex-based quadrilateral" items seemed to benefit from minimal directions in grade band 11.

When asked if they had difficulty using the computer, ELA students, in grade band 3, under minimal directions, said they had trouble with both "select text" and "reorder text" items. The ELA grade band 11 students also seemed to have some difficulty with the "reorder text" items. Since these difficulties were related to specific item types, the results suggest that there was uncertainty about how to perform the task, rather than uncertainty about using the computer itself. Mathematics students did not seem to have any problems using the computer.

*Research Question 7: Smarter currently intends to administer the passage first, and then administer the items one item at a time. Does this affect student performance?*

Smarter Balanced is interested in the possibility of administering items adaptively within a passage. This would require administering items sequentially so that the ability estimate could be updated after each item. Presenting items one at a time may take longer, and students may object to not knowing what is coming next. This question is designed to assess whether administering an item set takes longer when the items are presented sequentially and whether there is a difference in confusion or frustration level when students are presented a passage and all the items together or are presented a passage with the items then being presented one at a time. The item sets were not administered adaptively.

Two sets of items were created for a given test form. Both sets contained passages of equivalent length and difficulty as well as items of equivalent difficulty.[1] The first set in a form presented the passage with all the items together. The second set presented the passage with the items presented one at a time.

The forms were administered, within grade band, to different samples of students. Each sample contained both a general education group (Gen Ed) and a group that received English language accommodations (ELL) students. One sample was timed without thinking aloud during the administration. Each item set in these forms was separately timed. This sample provided timing information only. The second sample involved thinking aloud while responding to the questions and was not timed.

The primary questions of interest were:

1. Does presenting the items individually after the passage appear to take longer (timed condition)?
2. Does presenting the items individually after the passage increase the student's negative emotional states (e.g., frustration, confusion; think-aloud condition)?
3. Do students prefer one approach or another (think-aloud condition)?

The time it took to complete the sets when all items were presented together or one at a time varied by grade band and sample. For the grade band 3 and grade band 11 samples, timing differed little whether the items were presented at once or one at a time. However, for grade band 6, presenting the items one at a time took substantially longer for both the Gen Ed and ELL samples. While there is some variability between the ELL and the Gen Ed samples, the differences are not large and show the same pattern within grade band.

There appears to be slightly more *confusion* for both the Gen Ed and the ELL samples in grade band 3 when all the items are presented together. However, similar *frustration* levels were observed under the two formats for the grade band 3 students. Students working on the grade band 6 ELL sample showed similar patterns of frustration and confusion in both presentation formats. However, the Gen Ed grade band 6 students showed slightly more confusion when the items were presented one at a time.

---

[1] Comparable passage difficulty was achieved through the use of readability and lexile measures. Comparable item difficulty was achieved through DOK measures.

The grade band 6 students tended to score higher when the items were presented all at once (for both the Gen Ed students and the ELL students). The grade band 3 students showed similar results, regardless of sample or administration format. The grade band 11 Gen Ed students scored higher when the items were presented one at a time, while the grade band 11 ELL sample students scored higher when the items were presented altogether.

Both the ELL and Gen Ed grade band 3 students preferred to have the items presented one at a time. Grade band 11 students had a slight bias toward having the items presented one at a time. Conversely, grade band 6 students preferred to have the items presented together.

*Research Question 8: Smarter intends to present relatively long passages. Do longer passages reduce student engagement?*

Smarter Balanced is interested in using passages that are longer than those presently used. The Smarter Balanced recommended passage lengths are: for grades 3–5: 450–562 words for short passages and 563–750 words for long passages; for grades 6–8: 650–712 words for short passages and 713–950 words for long passages; and for high school, 800–825 words for short passages and 826–1100 words for long passages. There is concern that the longer passages may tax the processing abilities of ELL students and students with disabilities (SWD).

This question is designed to assess whether longer passages reduce student engagement, hamper the completion of the longer passages, or affect the depth of processing of the passage. Two sets of items were created. Both sets contained passages of equivalent difficulty with four items of equivalent difficulty attached to each passage. Both sets present the passage and all the items together. Each form contained a standard-length passage and an extended-length passage. The first set contained a passage of standard length. The second set contained a passage that is longer than standard length (extended-length, the length equivalent to that intended for use by Smarter Balanced).

The design was intended to compare the performance of two groups of students—ELL/SWD and Gen Ed students—across three grade bands: 3, 6, and 11. Twelve students took the forms. Of these, nine were grade band 3 Gen Ed students and one grade band 3 student was classified ELL/SWD. The single grade band 6 student was an ELL/SWD student. The two grade band 11 students were Gen Ed students.

All the ELL/SWD students were unaffected by the use of the longer passage. They were able to read the entire passage regardless of passage length and demonstrated that the longer passage was processed at a deep level. The ELL/SWD students also were not bored or distracted while reading either passage.

On the contrary, Gen Ed students did appear to be affected by the longer passage in grade bands 3 and 11. About 75 percent of the grade band 3 students and all of the grade band 11 students were affected by the use of the longer passage. Only 43 percent of the grade band 3 Gen Ed students and 50 percent of the grade band 11 Gen Ed students demonstrated a level of deep processing. Also, some percentage of the Gen Ed students were bored, regardless of the length of the passage

*Research Question 9: How long does it take for students to read through complex texts, performance tasks, etc.? Is timing affected by the way students are presented the passage and items?*

One way of making items more difficult is to increase their complexity. Complex items often take longer to solve or answer. In computer adaptive tests, added complexity may decrease the time a high ability student has to complete the test if the items are made more difficult through increased complexity. This potentially creates some fairness issues in an adaptive test if there is a time limit on the test. This question was designed to assess the time it takes for students to answer complex and simpler items. Complexity was defined as a function of the DOK demanded by the test question. It was hypothesized that more complex tasks would take more time.

Each ELA form had six items. These items varied in item complexity (simple or complex) and item format (SR, TE, or CR). The TE items were all "hot text" (HT) items. These items require the student to either highlight the text or drag the text to answer the item.

Forms were constructed in ELA at two grade bands: grade band 3–5 (referred to as grade band 3) and grade band 6 and 7 (referred to as grade band 6). Two forms were administered in grade band 3. One form was administered in grade band 6.

It was hypothesized that more complex items would take longer to complete than simpler items, but no evidence was found to support this hypothesis. SR items were answered in the shortest time. HT items took about one minute longer than SR items. CR items took the most time to answer, about 75 seconds longer then the hot text items.

*Research Question 10: Working mathematics problems on computer: Communicating mathematics on computer—feasibility of measuring student understanding of items for Claims 2–4 on computer.*

With paper tests some students write in their test books while working out mathematics problems. When mathematics items are presented on computer, scratch paper is often provided if students want to transfer the problem to paper and work it out there. Because scratch paper is often destroyed after an online testing session, the degree to which scratch paper is used is not known; neither is the importance of scratch paper in working out a problem (or potentially for use in scoring). This research question examines the need for paper when solving mathematics problems.

Each student was presented with three grade-appropriate items. The interviewer recorded whether the student made a comment, and the nature of the comment, while working the mathematics problems. The students first tried to work a problem without paper. Scratch paper was then offered to the student to rework the problem, if desired. The interviewer noted whether students chose to add anything additional and noted the nature of the addition (more text, equations, graphics). Note that there were only three comments for the third item in the lowest grade band, 3.

The general conclusion is that a subset of students benefit from being able to work mathematics problems on paper. This appears to be especially important when students are beginning to learn algebra concepts.

Grade band 3 students did not need paper to work the problems. However, in the grade band 6 and grade band 7 groups, 30–42 percent indicated they wanted to write an equation. In grade bands 6, 7, and 11, the additional information recorded on paper would have improved the response according to the rubric. Responses for specific items in grade bands 6 and 11 were improved by 15 percent of the students, and responses for all items in grade band 7 were improved when information on the scratch paper was taken into account. Improvement for this group ranged between 10 and 20 percent of the responses. ("Confused me, I didn't know how to write an equation." "Tried the keypad, but it wouldn't work." "It was much easier with paper.") This was supported by interviewer observations. About 5–10 percent of students in each grade band found the online system difficult to use, but few specifics were recorded.

*Research Question 11: Usability of equation editor tool—can students use the tool the way it is meant to be used?*

Although students begin to use technology at a very early age, it is prudent to verify that young students are able to use the assessment interface to be used during testing. This question sought to evaluate the ability of grades 3–5 students to use the equation editor tool to be included in the Smarter Balanced delivery system. Three mathematics items were presented to the students (*N*=33). The first item only required the student to copy his or her response. The second item was a simple mathematics item, and the third item was a more challenging mathematics item. The first item would demonstrate whether the student could use the equation editor tool. The second and third items would provide evidence of whether the ability to use the tool interacted with item difficulty.

Elementary students had some difficulty using the equation editor. Between 15 and 30 percent of the students indicated that they had difficulty using the equation editor. The examiner's assessment concurred that about 35 percent of students had difficulty using the equation editor and that about 50 percent of the students would get a given item correct.

*Research Question 12: Can students compare the size of a product to the size of one factor, on the basis of the size of the other factor, without performing the indicated multiplication?*

This question is designed to assess whether students with a strong understanding of fractions and the multiplication and division of fractions complete the items without performing the indicated multiplication. The task asked students to compare the size of a product to the size of one factor, on the basis of the size of the other factor, without performing the indicated multiplication. Also of interest was whether students who complete an item as intended (without using multiplication) spent less time on an item than those who did not. To investigate this question a single form was administered for grades 3–5.

There seemed to be little relationship between whether a student has a strong understanding of the multiplication and division of fractions and whether he or she used multiplication to solve the items. However, students who did not need to perform the multiplication completed the items in less time than students who had to perform the multiplication. While most students said they understood the questions, 70 percent had to use multiplication to solve them. Only about 40 percent of the students had a firm understanding of the multiplication/division of fractions, according to the interviewers.

*Research Question 13: Contextual glossaries are item-specific glossaries that provide a definition of a word that is targeted to, and appropriate for, the context in which the word is used in the item. Are these a fair and appropriate way to support students who need language support?*

This question addressed the efficacy of the use of contextual glossaries with non-native speakers when solving mathematics problems. Two sets of items were created that were parallel in difficulty. The first set of items contained no contextual glossaries with only single words translated. The second set of items contained contextual glossaries. The interviewer was asked to determine whether the student was having trouble understanding a word and whether the contextual glossary aided in the interpretation of the word or sentence.

Only three ELL students participated: one from grade 3 and two from grade 6.

The contextual glossaries appeared to be somewhat effective, but the impact was not always reflected in the score the student received for an item. The contextual glossaries appeared to be incomplete in that they did not include words the student needed. This limited the use of the glossaries in these situations. Interviewer's comments suggested that performance was improved when the students used the contextual glossaries.

*Research Question 14: Under what conditions does the use of text-to-speech (TTS) help students with lower reading ability focus on content in ELA and mathematics?*

TTS can provide access to an assessment for students with low reading ability. In order for this technology to be effective the language produced from the voice-pack must be clear enough to be understood. This is particularly true for non-native speakers of English.

Only students familiar with TTS were included in the study. Overall, 77 students used TTS at least once. Among them, 58 students were limited English proficient (LEP), 13 students had reading difficulties (IEP), and six were Gen Ed students.

In ELA four forms were administered with both high- and low-quality voice-packs. In mathematics, two forms were administered in grade bands 3 and 11. Only a single form was administered in grade band 6. The mathematics forms were only administered with high-quality voice-packs.

TTS improved access in ELA regardless of the quality of the voice-pack. Greater access was achieved when high-quality voice-packs were used. LEP students and students with reading difficulties tended to benefit more from the use of TTS. Using TTS with high-quality voice-packs improved focus on content in ELA. The use of TTS with low-quality voice-packs tended to distract students in ELA, whereas high-quality voice-packs did not. In mathematics, access was improved only for grade band 3 students. All Gen Ed, IEP, and grade band 6 LEP students found the high-quality voice-pack distracting. This was in part a function of trying to describe a table verbally.

## Introduction

Smarter Balanced has conducted cognitive laboratories to better understand how students solve items in different formats. A cognitive laboratory uses a think-aloud methodology in which students speak their thoughts while solving a test item. The interviewer follows a standardized protocol to elicit responses and record what a student says. While this one-on-one process is time consuming, the type of information elicited is often difficult to obtain by other means. Due to the nature of the process the sample sizes are often small; however, they are sufficient to detect large effects. In addition, because each student's comments are recorded, smaller, non-primary effects may be brought to light. Most protocols were developed at multiple grade bands (e.g., 3, 6, and 11). A grade band is the level of content for which the protocol is targeted.

What follows are in-depth analyses for each research question outlined in the executive summary. Because of the differences in the samples, study design, and questions asked, each research question result is presented separately. A summary of the findings for each research question is provided at the end of each research question section. Research questions have been organized into sections of similar content to improve integration of the material. Finally, a conclusions section appears at the end of the document. The overall demographics for the cognitive labs sample can be found in Appendix B.

## Processing Selected-Response (SR), Technology-Enhanced (TE), and

## Constructed-Response (CR) Items

*Research Question 1: Do mathematics multi-part selected-response (MPSR) items provide similar information about the depth of understanding by the test taker as do traditional constructed-response (CR) Items?*

An MPSR item has students select several examples of a correct response rather than just one, as in the typical SR item. The intention of this research question was to see whether the MPSR items provided depth of understanding similar to that of CR items. If effective, an MPSR item would be a more efficient way to measure the content measured by CR items. Also of interest was whether similar results would be obtained at different educational levels. To investigate these questions, forms were constructed at four grade bands: grades 3–4 (referred to as grade band 3), grades 6–7 (referred to as grade band 6), grades 7–8 (referred to as grade band 7), and grades 9–10 (referred to as grade band 11). Within a form, parallel items were constructed in both formats and presented to the same students. In the protocols, the MPSR and CR items were presented in random order.

Interviewers were asked to assess the highest level of DOK the student demonstrated during the think-aloud session. Table 1 (ELA) and Table 2 (mathematics) show the rubrics the interviewers used during this process.

Two hypotheses related to research question 1 were examined. The first hypothesis examined whether students who get full credit on MPSR items reveal, through their think-aloud sessions, greater understanding than those students who do not achieve full credit. The second hypothesis examined whether students who get full credit on MPSR items reveal understanding similar to that of students who get full credit on similarly challenging CR items measuring the same target.

Table 1. Depth of Knowledge Chart (ELA)

| DOK Level | Definition | Types of statements |
|---|---|---|
| 1 | Recall and Reproduction | 1. Recalls facts, details, and events<br><br>2. Uses word relationships (synonym/ antonym) to determine meaning<br><br>3. Recognizes or retrieves information from tables and charts |
| 2 | Basic Skills and Concepts | 1. Summarizes information<br><br>2. Identifies central ideas<br><br>3. Uses context to determine word meanings<br><br>4. Analyzes text structure and organization<br><br>5. Compares literary elements, facts, terms, or events |
| 3 | Strategic Thinking and Reasoning | 1. Uses supporting evidence to explain, generalize, or connect ideas<br><br>2. Analyzes or interprets author's craft (literary devices, viewpoint, potential bias) to critique a text<br><br>3. Develops a logical argument and cites evidence |

Table 2. Depth of Knowledge Chart (Mathematics)

| DOK Level | Definition | Types of statements |
|---|---|---|
| 1 | Recall and Reproduction | I remembered it.<br>We learned the answer in class.<br>I did what it said.<br>I recognized it. |
| 2 | Basic Skills and Concepts | 1. Any statement indicating putting two or more pieces of knowledge together<br>2. An statement indicating that they executed a sequence of steps that was not given to them<br>3. Any inference relating two different things<br>4. Expression of a hypothesis or guess about a relationship |
| 3 | Strategic Thinking and Reasoning | 1. Any statement indicating that they are applying abstract concepts to concrete phenomenon, e.g., "Both patterns reflect exponential growth"<br><br>2. Statements indicating that the students evaluated several different approaches to solving the problem, accompanied by the ability to explain why they selected the solution path they chose<br><br>3. Explanations of their choices or decisions using data and information from multiple sources to construct a coherent and logical argument |

## Results

Twenty students were administered the grade band 3 form, 37 students were administered the grade band 6 form, 31 students were administered the grade band 7 form, and 19 students were administered the grade band 11 form.

Table 3 presents the average DOK demonstrated level by students who received full credit on an item for each grade band/target. Table 4 shows the correspondence between the target labels and the full target description. Blank cells are the result of incomplete data, either in the score or in the demonstrated DOK. In most cases, the DOK the student demonstrated for the MPSR items either equals or exceeds the DOK demonstrated for the CR items. Interviewers commonly commented that the student did equally well on both item formats.

Table 3. Average DOK Demonstrated by Students Who Received Full Credit for Paired MPSR and CR Items Measuring the Same Target

| Grade Band | Target | Item Format | Avg. DOK |
|---|---|---|---|
| 3 | Geometric Measurement: Perimeters (J) | MPSR | 2.00 |
| 3 | Geometric Measurement: Perimeters (J) | CR | 1.50 |
| 3 | Reason with Shapes (K) | MPSR | 1.80 |
| 3 | Reason with Shapes (K) | CR | 1.67 |
| 6 | One Variable Equations (F) | MPSR | 1.57 |
| 6 | One Variable Equations (F) | CR | 1.60 |
| 6 | Analyze Proportional Relationships (A) | MPSR | |
| 6 | Analyze Proportional Relationships (A) | CR | 1.25 |
| 6 | Generate Equivalent Expressions (C) | MPSR | 2.00 |
| 6 | Generate Equivalent Expressions (C) | CR | 2.00 |
| 6 | Apply Arithmetic to Algebra (E) | MPSR | 1.60 |
| 6 | Apply Arithmetic to Algebra (E) | CR | 2.00 |
| 7 | Analyze Proportional Relationships (A) | MPSR | |
| 7 | Analyze Proportional Relationships (A) | CR | 1.83 |
| 7 | Generate Equivalent Expressions (C) | MPSR | 1.77 |
| 7 | Generate Equivalent Expressions (C) | CR | 2.00 |
| 7 | Solve Linear Equations (D) | MPSR | 2.00 |
| 7 | Solve Linear Equations (D) | CR | 1.80 |
| 11 | Equivalent Problem Solving (E) | MPSR | 2.33 |
| 11 | Equivalent Problem Solving (E) | CR | 1.75 |
| 11 | Graph Equations and Inequalities (J) | MPSR | |
| 11 | Graph Equations and Inequalities (J) | CR | 1.70 |
| 11 | Use of Functions (K) | MPSR | 2.10 |
| 11 | Use of Functions (K) | CR | 2.00 |

Table 4. Correspondence Between Target Label and the Full Target Description

| Target Label | Full Target Description |
|---|---|
| Geometric measurement: Perimeters | Geometric measurement: recognize perimeter as an attribute of plane figures and distinguish between linear and area measures |
| Reason with Shapes | Reason with shapes and their attributes |
| Place Value: Whole Numbers | Generalize place value understanding for multi-digit whole numbers |
| Converting Units of Measure | Solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit |
| Geometric measurement : Perimeters | Geometric measurement: recognize perimeter as an attribute of plane figures and distinguish between linear and area measures |
| One Variable Equations | Reason about and solve one-variable equations and inequalities |
| Apply Arithmetic to Algebra | Apply and extend previous understandings of arithmetic to algebraic expressions |
| Generate Equivalent Expressions | Use properties of operations to generate equivalent expressions |
| Analyze Proportional Relationships | Analyze proportional relationships and use them to solve real-world and mathematical problems |
| Solve Linear Equations | Analyze and solve linear equations and pairs of simultaneous linear equations |
| Equivalent Problem Solving | Write expressions in equivalent forms to solve problems |
| Graph Equations and Inequalities | Represent and solve equations and inequalities graphically |
| Use of Functions | Understand the concept of a function and use function notation |

The second hypothesis examined whether students who get full credit on the MPSR items reveal greater understanding of the material than those who do not obtain full credit. Table 5 presents these findings. In all cases those who receive full credit for an item showed greater understanding than those who did not receive full credit. The percentage understanding is also quite similar for the MPSR and CR items.

Table 5. Percentage of Students Who Appear to Understand the Material, by Item Type, Grade Band, and Whether Full Credit Was Received

| | Grade Band | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | | 6 | | 7 | | 11 | |
| Item | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit |
| MPSR1 | 20 | 50 | 17 | 89 | 38 | 78 | 64 | - |
| CR1 | 12 | 100 | 25 | 100 | 55 | - | 40 | 100 |
| MPSR2 | 0 | 57 | 29 | 100 | 42 | 83 | 45 | 100 |
| CR2 | 7 | 75 | 23 | 90 | 36 | 75 | 67 | 67 |
| MPSR3 | 0 | - | 10 | 67 | 48 | 100 | 33 | 75 |
| CR3 | 0 | - | 8 | 100 | 50 | 75 | 58 | 67 |

## Summary

This research question sought to address two hypotheses. The first hypothesis examined whether students who get full credit on MPSR items reveal, through their think-aloud sessions, greater understanding than those students who do not achieve full credit. The second hypothesis examined whether students who get full credit on MPSR items reveal depth of understanding similar to that of students who get full credit on similarly challenging CR items measuring the same target.

In most cases, the DOK the student demonstrated for the MPSR items either equaled or exceeded the DOK demonstrated for the CR items.  Students who got full credit on the MPSR items also revealed greater understanding of the material than those who did not obtain full credit. The percentage of students understanding the material was also quite similar for the MPSR and CR items. A typical interviewer comment was, "based on the accuracy of the student's responses to both types of items, it appears that item type is not a factor in determining how well the students respond[s]."

*Research Question 2: Under what conditions do specific types of TE items (and SR items) approach the depth of knowledge (DOK) of a written constructed response in ELA and mathematics?*

The question was designed to assess whether different types of TE items approach the DOK of CR items for specific content claim/targets and DOK levels. SR items were also included, where available, as a comparison item format. Comparisons were examined for specific TE item types at specific DOK levels for specific content claims/targets (see Appendix A for a full description of the claims and targets). Where possible, parallel items were created in each item format at the same DOK level and content claim/target; however, some combinations were not available. In ELA, items in the different formats were administered for most item type/content target/DOK combinations. In mathematics, however, some item formats were not administered for all claim/target/DOK conditions and some data were incomplete. This limited the comparisons that could be made. Four items in one of the three formats (MPSR, TE, and CR) appeared in each form. Multiple forms were administered, each to a different sample of students. It was hypothesized that students responding to items of a specific TE type would reveal that they were using thought processes consistent with a specific DOK level for items measuring a specific target.

Forms were constructed in ELA at five grade bands: grade 3 (referred to as grade band 3), grades 4–5 (referred to as grade band 4), grades 6–7 (referred to as grade band 6), grades 7–8 (referred to as grade band 7), and grade 11 (referred to as grade band 11). In mathematics, forms were constructed at four grade bands: grades 3–4 (referred to as grade band 3), grades 4–5 (referred to as grade band 4), grades 6–7 (referred to as grade band 6), and grade 11 (referred to as grade band 11). Note that the grade band relates to the level of the material in the assessment and not necessarily the grade of the students to which the assessment is administered. A single form was administered in each grade band. This was a between-subjects design in which different item types were administered to different students. For this question, the comments presented are made by the interviewer, as opposed to the student, due to the nature of the information being captured (e.g., DOK level demonstrated).

## Results

Table 6 shows the sample sizes within a grade band by item format across item types and content area. The ELA forms tended to have been administered to larger samples than were the mathematics forms.

Table 6. Sample Sizes Within Grade Band, by Content Area and Item Type

| Content | Item Format | Grade Band | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 4 | 6 | 7 | 11 |
| ELA | MPSR | 18 | 16 | 13 | 8 | 6 |
| ELA | TE | 12 | 14 | 10 | 8 | 14 |
| ELA | CR | 14 | 13 | 13 | 15 | 10 |
| Mathematics | MPSR | 7 | 6 | 23 | - | 10 |
| Mathematics | TE | 7 | 4 | 13 | - | 3 |
| Mathematics | CR | 4 | 11 | 8 | - | 3 |

Tables 7a (ELA) and 7b (Mathematics) list the percentage of students whose thought processes were consistent with the DOK level of the items for the respective content areas. For each TE item type, the percentage of students who demonstrated thought processes consistent with the grade band/content claim and target/DOK was recorded. MPSR and CR items were matched to the same grade band/content claim and target/DOK levels. The primary comparison of interest is between the TE and CR formats.

For ELA, students demonstrated a higher DOK level for most of the TE item types than for the matched CR items. ("Well thought out. Uses evidence she feels supports the main idea of the item.") Two exceptions were two targets in the "select text" item type: "justifying interpretations" (grade band 6) and "analyzing the figurative" (grade band 11). A pattern similar to that of the TE item types was observed for the matched MPSR items versus the CR items.

Table 7a. Percentage of Students Demonstrating That They Are Using Thought Processes at the Specified DOK level, by Item Type, Claim, Target, and DOK Level (ELA)

| TE Item Type | Grade Band | Target | Claim | DOK | % of Students With Consistent Thought Process | | |
|---|---|---|---|---|---|---|---|
| | | | | | TE | MPSR | CR |
| Drag and Drop (Tiling) | 6 | Justifying interpretations | 1 | 3 | 63 | 78 | 40 |
| Drag and Drop (Tiling) | 7 | Writing or revising strategies | 2 | 2 | 100 | 80 | 61 |
| Reorder Text | 3 | Writing or revise strategies | 2 | 2 | 81 | 69 | 54 |
| Reorder Text | 6 | Organizing ideas | 2 | 2 | 60 | | |
| Select Text | 6 | Justify interpretations | 1 | 2 | 33 | 50 | 60 |
| Select Text | 7 | Identifying text to support inferences | 1 | 2 | 94 | 79 | 64 |
| Select Text | 7 | Writing or revising strategies | 2 | 2 | 100 | 80 | 61 |
| Select Text | 11 | Citing to support inferences | 1 | 2 | 72 | 82 | 69 |
| Select Text | 11 | Analyzing the figurative | 1 | 2 | 33 | 50 | 55 |

For mathematics, the pattern is less clear. The TE item types that yielded a higher percentage of students demonstrating thought processes consistent with the DOK level included:

- "placing points" for fractions, claim 1, DOK 2 in grade band 3 ("This student had a thorough understanding of these fractions and how they related to the number line. He thoroughly and accurately placed each fraction and explained how/why using various steps.")
- "single lines" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "tiling" for fractions, claim 1, DOK 2 in grade band 3 ("This student clearly understood and explained how to solve this item using multiple methods. He used multiple steps to solve each item.")
- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11 ("Student indicated use of multiple steps and solved correctly.")
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 4, DOK 3 in grade band 4

For the following TE item types, the percentages of students demonstrating thought processes consistent with the DOK level were equal for the TE and CR formats.

- "select and order" for fractions, claim 1, DOK 2 in grade band 3
- "select and order" for fractions, claim 1, DOK 2 in grade band 6
- "selecting points" for fractions, claim 1, DOK 2 in grade band 3
- "single lines" for everyday math problems, claim 2, DOK 2 in grade band 11

The TE item types for which the matched CR items yielded a higher percentage of students who demonstrate consistent thought processes included:

- "select and order" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6
- "tiling" for everyday mathematic problems, claim 4, DOK 3 in grade band 4
- "tiling" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6 ("The student was able to explain his answer in multiple steps and with a clear understanding of the distributive property.")
- "tiling" for everyday mathematic problems, claim 2, DOK 3 in grade band 11
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4 ("This student understood right angles. She also understood that she had to name a similarity and a difference.")

Table 7b. Percentage of Students Demonstrating That They Are Using Thought Processes at the Specified DOK Level, by Item Type, Claim, Target, and DOK Level (Mathematics)

| TE Item Type | Grade Band | Target | Claim | DOK | % of Students With Consistent Thought Process | | |
|---|---|---|---|---|---|---|---|
| | | | | | TE | MPSR | CR |
| Placing Points | 3 | Fractions | 1 | 2 | 50 | 53 | 0 |
| Select and Order | 3 | Fractions | 1 | 2 | 0 | 53 | 0 |
| Select and Order | 6 | Apply arithmetic to algebraic expressions | 1 | 2 | 40 | 67 | 79 |
| Select and Order | 6 | Everyday math problems | 2 | 3 | 50 | | |
| Selecting Points | 3 | Fractions | 1 | 2 | 0 | 53 | 0 |
| Single Lines | 11 | Equations and inequalities | 1 | 2 | 50 | 82 | 42 |
| Single Lines | 11 | Everyday math problems | 2 | 2 | 100 | | 100 |
| Tiling | 3 | Fractions as numbers | 1 | 2 | 71 | 53 | 0 |
| Tiling | 4 | Everyday math problems | 4 | 3 | 0 | 33 | 52 |
| Tiling | 6 | Apply arithmetic to algebraic expressions | 1 | 2 | 60 | 67 | 79 |
| Tiling | 11 | Equations and inequalities | 1 | 2 | 50 | 82 | 42 |
| Tiling | 11 | Everyday math problems | 2 | 3 | 0 | 39 | 50 |
| Vertex-Based Quadrilaterals | 4 | Lines, angles, and shapes | 4 | 3 | 67 | 33 | 52 |
| Vertex-Based Quadrilaterals | 4 | Lines, angles, and shapes | 1 | 2 | 33 | 83 | 72 |

Also of interest was how students performed on these item types. Since not all items are 1-point items, the percentage obtaining the maximum score was used. Table 8a presents this information for ELA; Table 8b presents this information for mathematics. In ELA, the pattern is very similar to the consistency of thought process table. The same TE item types had higher percentages of students with maximum scores than did the CR items, with the exception of the "select text" items for the "writing or revising strategies" target (grade band 7), and the "citing to support inferences" target (grade band 11).

For the MPSR items in ELA, the percentage receiving the maximum score was higher than both the CR and TE formats for the following "select text" items:

- "select text" for justifying interpretations, claim 1, DOK 2 in grade band 6
- "select text" for citing to support inferences, claim 1, DOK 2 in grade band 11
- "select text" for analyzing the figurative, claim 1, DOK 2 in grade band 11

Table 8a. Percentage of Students Receiving Full Credit for an Item (ELA)

| TE Type | Grade Band | Target | Claim | DOK | % of Students With Maximum Score | | |
| | | | | | TE | MPSR | CR |
|---|---|---|---|---|---|---|---|
| Drag and Drop (Tiling) | 6 | Justifying interpretations | 1 | 3 | 80 | 67 | 18 |
| Drag and Drop (Tiling) | 7 | Writing or revising strategies | 2 | 2 | 67 | 22 | 47 |
| Reorder Text | 3 | Writing or revise strategies | 2 | 2 | 64 | 0 | 44 |
| Reorder Text | 6 | Organizing ideas | 2 | 2 | 12 | | |
| Select Text | 6 | Justifying interpretations | 1 | 2 | 10 | 70 | 25 |
| Select Text | 7 | Identifying text to support inferences | 1 | 2 | 77 | 19 | 41 |
| Select Text | 7 | Writing or revising strategies | 2 | 2 | 0 | 22 | 47 |
| Select Text | 11 | Citing to support inferences | 1 | 2 | 22 | 67 | 40 |
| Select Text | 11 | Analyzing the figurative | 1 | 2 | 8 | 46 | 31 |

In mathematics, the TE items for which a higher percentage of students received the maximum possible score included only:

- "single lines" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4

For the MPSR items in mathematics, the percentage receiving the maximum score was higher than both the CR and TE formats for the following items:

- "placing points" for fractions, claim 1, DOK 2 in grade band 3
- "select and order" for fractions, claim 1, DOK 2 in grade band 3
- "selecting points" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for everyday math problems, claim 4, DOK 3 in grade band 4
- "vertex-base quadrilaterals" for fraction equivalence and ordering, claim 1, DOK 2 in grade band 3

In other cases the percentage receiving the maximum score was lower than for the comparable CR items. It should be noted that the percentage receiving the maximum scores was generally low in mathematics.

Table 8b. Percentage of Students Receiving Full Credit for an Item (Mathematics)

| TE Type | Grade Band | Target | Claim | DOK | TE | MPSR | CR |
|---|---|---|---|---|---|---|---|
| | | | | | % of Students With Maximum Score | | |
| Placing Points | 3 | Fractions | 1 | 2 | 17 | 35 | 25 |
| Select and Order | 3 | Fractions | 1 | 2 | 14 | 35 | 25 |
| Select and Order | 6 | Everyday math problems | 2 | 3 | 0 | | |
| Select and Order | 6 | Apply arithmetic to algebraic expressions | 1 | 2 | 0 | 21 | 44 |
| Selecting Points | 3 | Fractions | 1 | 2 | 14 | 35 | 25 |
| Single Lines | 11 | Everyday math problems | 2 | 2 | 0 | | 80 |
| Single Lines | 11 | Equations and inequalities | 1 | 2 | 33 | | 27 |
| Tiling | 3 | Fractions | 1 | 2 | 33 | 35 | 25 |
| Tiling | 4 | Everyday math problems | 4 | 3 | 0 | 25 | 5 |
| Tiling | 6 | Apply arithmetic to algebraic expressions | 1 | 2 | 31 | 21 | 44 |
| Tiling | 11 | Everyday math problems | 2 | 3 | 33 | 16 | 40 |
| Tiling | 11 | Equations and inequalities | 1 | 2 | 33 | 0 | 27 |
| Vertex-Based Quadrilaterals | 3 | Fraction equivalence and ordering | 1 | 2 | 21 | 35 | 25 |

| Vertex-Based Quadrilaterals | 4 | Lines, angles, and shapes | 4 | 3 | 0 | 25 | 5 |
|---|---|---|---|---|---|---|---|
| Vertex-Based Quadrilaterals | 4 | Lines, angles, and shapes | 1 | 2 | 67 | 50 | 0 |

## Summary

For ELA, students demonstrated a higher DOK level for most of the TE item types than for the matched CR items. Two exceptions were two targets in the "select text" item type, "justifying interpretations" (grade band 6) and "analyzing the figurative" (grade band 11). A similar pattern was observed for the matched MPSR items versus the CR items.

Regarding student performance on the ELA items, the pattern of results is very similar to the consistency of thought process table. The same TE item types had higher percentages than did the CR items, with the exception of the "select text" items for the "writing or revising strategies" target (grade band 7) and the "citing to support inferences" target (grade band 11).

For the MPSR items in ELA, the percentage receiving the maximum score was higher than both the CR and TE formats for the following "select text" items:

- "select text" for justifying interpretations, claim 1, DOK 2 in grade band 6
- "select text" for citing to support inferences, claim 1, DOK 2 in grade band 11
- "select text" for analyzing the figurative, claim 1, DOK 2 in grade band 11

For mathematics, the results were more varied. The TE item types that showed a higher percentage of students demonstrating consistent thought process with the DOK level included:

- "placing points" for fractions, claim 1, DOK 2 in grade band 3
- "single lines" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "tiling" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11 ("Student indicated use of multiple steps and solved correctly.")
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 4, DOK 3 in grade band 4

Places where equal percentages were observed for the TE and CR formats included:

- "select and order" for fractions, claim 1, DOK 2 in grade band 3
- "select and order" for fractions, claim 1, DOK 2 in grade band 6
- "selecting points" for fractions, claim 1, DOK 2 in grade band 3
- "single lines" for everyday math problems, claim 2, DOK 2 in grade band 11

Item types where the CR items had a higher percentage of consistent thought processes included:

- "select and order" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6
- "tiling" for everyday mathematic problems, claim 4, DOK 3 in grade band 4

- "tiling" for apply arithmetic to algebraic expressions, claim 1, DOK 2 in grade band 6
- "tiling" for everyday mathematic problems, claim 2, DOK 3 in grade band 11
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4

The TE item types where a higher percentage of students received full credit included only:

- "tiling" for equations and inequalities, claim 1, DOK 2 in grade band 11
- "vertex-base quadrilaterals" for lines, angles, and shapes, claim 1, DOK 2 in grade band 4

For the MPSR items in mathematics, the percentage receiving the maximum score was higher than both the CR and TE formats for the following items:

- "placing points" for fractions, claim 1, DOK 2 in grade band 3
- "select and order" for fractions, claim 1, DOK 2 in grade band 3
- "selecting points" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for fractions, claim 1, DOK 2 in grade band 3
- "tiling" for everyday math problems, claim 4, DOK 3 in grade band 4
- "vertex-base quadrilaterals" for fraction equivalence and ordering, claim 1, DOK 2 in grade band 3

In other cases, the percentage receiving full credit for the MPSR items was lower than for the comparable CR items. It should be noted that the percentage receiving full credit was generally low in mathematics for all three item formats.

*Research Question 3: For multi-part selected response (MPSR) items where students may select more than one answer choice, which wording best indicates to the student that he or she is allowed to select more than one option? For multipart (e.g., YES/NO) dichotomous choice items, do students know that they need to answer each part?*

Smarter Balanced sought to investigate whether students might become confused by MPSR items in mathematics and perhaps not complete the entire item. In order to investigate this, items were constructed with different amounts of labeling. Labeling is the identification of the parts of the problem with indicators such as "a," "b," "c" or "1," "2," "3." A "labeled" and a non-labeled" condition were investigated. An example of items in the labeled and unlabeled format is presented below (Exhibit 1).

This question is designed to assess whether labeling or not labeling an MPSR mathematics item produces a difference in performance. Results are reported in five grade bands. The five grade bands are designated as grade band 3 (which includes form difficulty levels 3 and 4), grade band 4 (which includes form difficulty levels 4 and 5), grade band 6 (which includes form difficulty levels 6 and 7), grade band 7 (which includes form difficulty levels 7 and 8), and grade band 11 (which includes form difficulty level 11). Each form contains one MPSR item followed by one CR item. The labeled and non-labeled items appeared in different forms of the test and thus were taken by different students.

Exhibit 1. Example of a Labeled Item

Marcus has 36 marbles. He is putting an equal number of marbles into 4 bags.

Indicate whether each equation could be used to find the number of marbles Marcus puts in each bag.

1.  $36 \times 4 = \boxed{\phantom{00}}$    ○ Yes   ○ No

2.  $36 \div 4 = \boxed{\phantom{00}}$    ○ Yes   ○ No

3.  $4 \times \boxed{\phantom{00}} = 36$    ○ Yes   ○ No

4.  $4 \div \boxed{\phantom{00}} = 36$    ○ Yes   ○ No

Example of an Unlabeled Item

Marcus has 36 marbles. He is putting an equal number of marbles into 4 bags.

Indicate whether each equation could be used to find the number of marbles Marcus puts in each bag.

$36 \times 4 = \boxed{\phantom{00}}$    ○ Yes   ○ No

$36 \div 4 = \boxed{\phantom{00}}$    ○ Yes   ○ No

$4 \times \boxed{\phantom{00}} = 36$    ○ Yes   ○ No

$4 \div \boxed{\phantom{00}} = 36$    ○ Yes   ○ No

## Results

Ninety-six students were administered the grade band 3 forms, 66 students were administered the grade band 4 forms, 133 students were administered the grade band 6 forms, 33 students were administered the grade band 7 forms, and 85 students were administered the grade band 11 forms.

Table 9 shows the percentage of students receiving full credit on the items by grade band and labeling condition. For grade bands 3, 4, 6, and 11 little difference between the labeled and non-labeled conditions is observed. However, in grade band 7 a higher percentage of students received full credit in the non-labeled format.

Table 9. Percentage of Students Receiving Full Credit, by Grade Band and Labeling Condition.

| | Grade Band | | | | |
|---|---|---|---|---|---|
| Condition | 3 | 4 | 6 | 7 | 11 |
| Non-Labeled | 32 | 32 | 20 | 62 | 16 |
| Labeled | 29 | 31 | 18 | 34 | 9 |

Table 10 shows whether the students understood the instructions under the different item labeling conditions. Up through grade band 6 the type of instructions received seemed to have little impact on the understanding of the instructions. However, in grade bands 7 and 11 a higher percentage of students tended not to understand the instructions when the items were labeled. The interviewers commented that "Student did not have a complete understanding of instructions" and "He said he understood, however, he only selected one bubble."

Table 10. Percentage Understanding the Instructions, by Grade Band and Labeling Condition

| | Grade Band | | | | |
|---|---|---|---|---|---|
| Condition | 3 | 4 | 6 | 7 | 11 |
| Non-Labeled | 63 | 83 | 93 | 97 | 84 |
| Labeled | 78 | 93 | 93 | 69 | 61 |

Table 11 shows the percentage of students who made comments about not understanding the instructions. Grade bands 3 and 11 had more comments about not understanding the instructions than the other grade bands, but the pattern was similar for labeled and non-labeled items. However, in grade band 7, non-labeled items generally received no comment, with labeled items receiving

more comments. This is consistent with a lower percentage of grade band 7 students understanding the instructions in the "labeled" condition.

Table 11. Did the Student Make Comments About not Understanding the Instructions (Percentage Making Comments)?

| Condition | Grade Band | | | | |
|---|---|---|---|---|---|
| | 3 | 4 | 6 | 7 | 11 |
| Non-Labeled | 34 | 17 | 15 | 3 | 33 |
| Labeled | 32 | 26 | 8 | 38 | 41 |

## Summary

Even though the labeling of MPSR items was intended to clarify the mathematic tasks for the students, in many cases it actually seemed to confuse the students. Little difference was observed between the labeled and non-labeled items in the lower grade bands (grade bands 3–6). However, students in grade band 7 tended to score higher with non-labeled items. Also, grade band 7 and 11 students tended to be confused by the labeling. In addition, the labeled items tended to receive more comments related to not understanding the instructions. The interviewer confirmed this, suggesting that the grade band 7 and 11 students better understood the instructions in the non-labeled condition than in the labeled condition.

*Research Question 4: Does the ability to move one or more sentences to different positions provide evidence of students' ability to revise text appropriately in the consideration of chronology, coherence, transitions, or the author's craft?*

Smarter Balanced is considering using items that have students reorder sentences to measure an editing/revising standard. Claim 2 of the standards states that students should be able to revise one or more paragraphs demonstrating specific narrative strategies (use of dialogue, sensory or concrete details, description), chronology, appropriate transitional strategies for coherence, or authors' craft appropriate to purpose (closure, detailing characters, plot, setting, or an event).

This question was designed to assess whether students' movement of one or more sentences to different positions provides evidence of students' ability to demonstrate consideration of chronology, coherence, transitions, or author's craft. Six ELA items were included in a test form. The forms were administered to five students: two in grade 5, two in grade 6, and one in grade 10. Because there is little difference in the pattern of responses and because the sample sizes are small, the results will be reported for the sample as a whole.

## Results

It was hypothesized that students who do well on these items would recognize the need to revise for chronology, coherence, transitions, or author's craft. Table 12 shows the percentage of students who recognize the need to revise for chronology, coherence, transitions, or author's craft for students who performed well on the items and those who performed poorly. The results show that students who performed well are more likely to consider chronology, coherence, transitions, or author's craft in their revisions than students who do not. Among the four writing skills examined, author's craft was considered less often than the other three writing skills.

Table 12. Percentage of Students Considering Targeted Writing Skills When Revising, by Those Students Who Performed Well and Those Who Performed Poorly

| Characteristic | Students Who Perform Well | Students Who Perform Poorly |
|---|---|---|
| Chronology | 100% | 33% |
| Coherence | 100% | 33% |
| Transitions | 100% | 33% |
| Author's Craft | 50% | 0 |

Also of interest was whether students referenced organization, coherence, transitions, or author's craft when moving sentences. Table 13 shows the percentage of students who considered each of the targeted writing skills relative to the number of appropriate and inappropriate sentence moves. The results suggest that students who make more appropriate sentence moves (and fewer inappropriate sentence moves) are more likely to consider the writing skills of chronology, coherence, and transitions; however, the pattern is less clear for consideration of author's craft.

Table 13. Percentage of Students Who Considered Chronology, Coherence, Transitions, and Author's Craft at Each Number of Appropriate and Inappropriate Sentence Moves

| % Students Who Recognized Need For | N Appropriate Sentences Moved | | | | | | | | N Inappropriate Sentences Moved | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Chronology | 38 | 50 | 50 | 67 | 75 | | 100 | 100 | 100 | 100 | 40 | 50 | 33 | 0 | | 0 |
| Coherence | 38 | 43 | 67 | 67 | 75 | | 100 | 100 | 100 | 100 | 40 | 50 | 33 | 0 | | 0 |
| Transitions | 38 | 50 | 50 | 67 | 75 | | 100 | 100 | 100 | 100 | 40 | 50 | 33 | 0 | | 0 |
| Anthor's craft | 13 | 13 | 0 | 33 | 25 | | 67 | 0 | 29 | 17 | 20 | 50 | 33 | 0 | | 0 |

Table 14 shows the percentage of students who considered chronology, coherence, transitions, and author's craft when answering the items as observed by the interviewers. Students did express consideration of chronology ("I moved the first sentence because it goes at the top," "This seems to be in order," "This should be the second to last sentence"); coherence ("This seems like something you'd say," "I don't need to take out more phrases, it sounds OK," "I removed the two sentences because they did not make sense and were irrelevant"); and transitions ("This would sound better here") when answering the questions; however, fewer took author's craft ("I think there is a flow to the story," "Some sentences are awkward and need to be moved") into account when answering these questions.

Table 14. Percentage of Students Who Considered Chronology, Coherence, Transitions, and Author's Craft When Answering, Across Items

| Writing Skills | Chronology | Coherence | Transitions | Author's Craft |
|---|---|---|---|---|
| Percentage | 68 | 68 | 57 | 18 |

## Summary

Students who performed well on the items were more likely to consider the targeted writing skills (chronology, coherence, transitions, and author's craft) when answering the questions. Also, students who made appropriate sentence moves were more likely to consider the targeted writing skills than those who made inappropriate sentence moves. A high percentage of students considered chronology, coherence, and transitions; however, they were less likely to consider author's craft.

*Research Question 5: Do students who construct text reveal more understanding of targeted writing skills than students who manipulate writing through the manipulation of text (MT) tasks?*

Many believe that the best way to measure writing is to have students write. However, in a testing environment, it is often difficult to adequately sample the writing content domain with an assessment composed exclusively of CR items. There is an ongoing effort to find items that are efficient, but that can adequately measure the components of the writing domain, thus allowing for a broader selection and greater number of items to be delivered. Examples of the types of questions used can be seen in Exhibit 2.  This question examined whether students responding to MT tasks would demonstrate understanding of the targeted writing skills comparable to the understanding demonstrated for matched CR tasks.

Four pairs of ELA items were developed. Each pair contained one MT item and one CR version of the same item. Two forms were created, with each form containing a single version of an item. Each form contained two MT items and two CR items. The MT items were almost exclusively "select and order" items, though two of the items - one in grade band 3 and one in grade band 11 - were "reorder text" items.  All items assessed claim 1, target 1.

Forms were constructed in ELA at three grade bands: grades 3–5 (referred to as grade band 3), grades 6 and 7 (referred to as grade band 6), and grades 10 and 11 (referred to as grade band 11). In grade band 3, two forms were administered; in grade bands 6 and 11, only a single form was administered. All forms assessed claim 1, target 1.

The sample consisted of seven students in grade band 3, two students in grade band 6, and one student in grade band 11.

Exhibit 2. Sample Items Used in this Research Question

<u>Stem</u>

A student wrote the first draft of a story about a girl who eats nine berries for an afternoon snack every day. Read the story. Then complete the task that follows.

> Every day after school, Kim eats nine red, juicy raspberries. One day, Kim sits down at the big kitchen table and has a surprise. She notices that one of her berries is missing! "[ ]," she says.
> "I counted nine just a minute ago," Dad says.
> "[ ]," Kim says. "[ ]."
> Kim begins her search in the garage. "[ ]?" Kim asks.

**Dialogue**
Oh no! There are only eight raspberries in my bowl

I wonder what happened to the ninth berry

Grandma, why are your mouth and lips red

It looks like I have a mystery to solve

Revise the story to include dialogue that introduces the plot. Place each piece of dialogue in the correct place in the story.

The dialogue will go in the brackets.

<u>CR Prompt</u>

A student wrote the first draft of a story about a girl who eats nine berries for an afternoon snack every day. Read the story. Then complete the task that follows.

---

Every day after school, Kim eats nine red, juicy raspberries. One day, Kim sits down at the big kitchen table and has a surprise. She notices that one of her berries is missing!

Her dad had counted nine just a few minutes ago.

Kim knew she had a mystery to solve.

Kim began her search in the garage. She found her grandmother in the garage with bright red lips.

---

Revise the story to include dialogue. Use dialogue to introduce the plot.
Type your response in the space provided.

Table 15. Targeted Writing Skills with Examples of Representative Statements

| Target | Types of Statements |
|---|---|
| Chronology | - I knew it was telling a story, so looked for the beginning then moved the rest around to make sense.<br>- I knew what the end was, so worked backwards from there.<br>- I knew the youngest son went last, so put him at the end, then put the two older ones before him. Then picked the beginning and put it first.<br>- Some spots didn't sound quite right, so added the sentences in.<br>- Read the sentences, then looked for related sentences in the passage that they'd go with.<br>- I used transitions to cue position of sentences.<br>- I need to revise the order of the sentences so that they more clearly support the main idea of the article. I do not need to move the first or last sentence. |
| Coherence | - Sentence is like a preview of the rest of the essay, so it should go first.<br>- This sentence sounds professional and it also connects to the facts that follow. This is the best thesis statement.<br>- This sentence wraps up the author's argument/point of view and finishes the essay by restating the main point.<br>- The conclusion often just rephrases the thesis, which this sentence does, but it also talks about other things from the passage, so it should be the conclusion.<br>- I have to choose the two sentences that shouldn't be part of the paragraph.<br>- I have to take the sentence at the top and drag it to best spot in the paragraph below. |
| Transitions | - The word "next" tells him it comes after something else.<br>- The word "first" is a clue that it goes at the beginning.<br>- "Finally" usually tells you you're at the end.<br>- A transition like "therefore" at the start of a sentence connects it to the sentence before. They have the same topic but this one comes second.<br>- I have to use transitions words to make the paragraph clearer.<br>- I looked at the transition words to see what should come before them, then put in a sentence if needed. |

| Author's Craft | - I found the parts that didn't give me a really clear picture in her mind and changed them.<br>- I looked for the parts that weren't as descriptive as the rest and made them more descriptive.<br>- I looked for the parts that sounded a little boring and made them more exciting.<br>- I read the topic sentences and looked for the sentence that didn't go with it.<br>- If a sentence makes the argument weaker, then it should be taken out, so these two need to be removed. |
|---|---|

## Results

It was hypothesized that student think-alouds on MT items would reference the appropriate writing skills reflected in the assessment target at a level comparable with CR items. Table 16 shows the percentage of students who referenced the targeted writing skills, by item format and grade band. In grade band 3, chronology was more likely to be considered during revision when the item format was MT ("First, next, last order of events") than when the item format was CR ("Historically probably comes first, having trouble ending story"). Similar patterns, but less pronounced, were seen with coherence, transitions ("This is a cause…as a result (an effect) should be here"), and author's craft. Grade band 3 students only considered author's craft during revision for about one-third of the items regardless of item format. Grade band 6 students always considered chronology and coherence during revision, but transitions and author's craft were only considered about half the time. In grade band 11 chronology, coherence, and transitions were always considered in both formats. Author's craft was only considered about half the time in the CR format and not mentioned at all in the MT format. One interviewer commented, "Student made no comment about author's craft."

Table 16. Percentage of Items in Which Students Considered Target Characteristics When Responding to the Item, by Item Format

| Target Characteristics | Item Format | Grade Band | | |
|---|---|---|---|---|
| | | 3 | 6 | 11 |
| Chronology | CR | 31 | 100 | 100 |
| | MT | 94 | 100 | 100 |
| Coherence | CR | 63 | 100 | 100 |
| | MT | 75 | 100 | 100 |
| Transitions | CR | 44 | 50 | 100 |
| | MT | 69 | 50 | 100 |
| Author's Craft | CR | 31 | 50 | 50 |

| | MT | 43 | 100 | 0 |
|---|---|---|---|---|

Table 17 shows the counts for item scores received for the two item formats, by grade band. Comparable scores were achieved for the two item formats.

Table 17. What Score (Across Items) Would the Student Receive on this Type of Item?

| | | Grade Band | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 3 | | | 6 | | | 11 | | |
| | Item Format | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| | CR | 8 | 5 | 3 | 0 | 0 | 2 | 0 | 0 | 2 |
| | MT | 7 | 6 | 2 | 1 | 0 | 1 | 0 | 2 | 0 |

Table 18 provides information about whether the students who construct text through writing reveal comparable or greater understanding of targeted writing skills than students who manipulate text. The grade band 3 and grade band 6 students were either more effective in applying the targeted writing skills when the items were in a MT format or no differences were observed in effectiveness between item formats. For the grade band 11 students the results were mixed, but students tended to be more effective in applying the targeted writing skills in the CR format, particularly for transitions and author's craft.

Table 18. Effectiveness of Applying Targeted Writing Skills by Item Format (Percentage of Students as Assessed by Interviewer)

| | Grade Band | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 | | | 6 | | | 11 | | |
| Target Characteristics | MT More Effective | No Difference | CR More Effective | MT More Effective | No Difference | CR More Effective | MT More Effective | No Difference | CR More Effective |
| Chronology | 38 | 63 | 0 | 0 | 100 | 0 | 0 | 100 | 0 |
| Coherence | 38 | 63 | 0 | 0 | 100 | 0 | 100 | 0 | 0 |
| Transitions | 25 | 75 | 0 | 50 | 50 | 0 | 0 | 0 | 100 |
| Author's Craft | 38 | 63 | 0 | 50 | 50 | 0 | 0 | 0 | 100 |

## Summary

The results showed that the targeted writing skills are considered by students who manipulate text at a level comparable to (or greater than) that encountered when they are constructing text. The grade band 3 and 6 students showed comparable (or greater) levels of understanding when the items were in an MT format. For the grade band 11 students the results were mixed, but students tended to be

more effective in applying the targeted writing skills in the CR format, particularly for transitions and author's craft. Score distributions were comparable for MT and CR item formats.

*Research Question 6: Do different types of directions (minimal, concise, or extensive) have an effect on the performance of technology enhanced (TE) items in ELA and Mathematics?*

The optimal amount of direction that should be given to students for some item types is unclear. With minimal directions students may not know how to approach the item; with extensive directions students may be distracted or slowed to a point where the item becomes inefficient. This may be particularly true with elementary school students, who may take longer to process text. This question examined these issues for ELA and mathematics items. Three types of directions (minimal, concise, and extensive) were examined for different item types.

Forms were constructed in ELA at five grade bands: grade 3 (referred to as grade band 3), grades 4 and 5 (referred to as grade band 4), grades 6 and 7 (referred to as grade band 6), grades 7 and 8 (referred to as grade band 7), and grade 11 (referred to as grade band 11) with a single form administered in each grade band. In mathematics, forms were constructed at four grade bands: grades 3 and 4 (referred to as grade band 3), grades 4 and 5 (referred to as grade band 4), grades 6 and 7 (referred to as grade band 6), and grade 11 (referred to as grade band 11).

Parallel items were created with minimal, concise, or extensive directions in ELA and for most item types in mathematics. However, not all direction types appeared with all item types in all grades in mathematics. Four items in one of the three formats (SR, TE, and CR) appeared in each mathematics form. Two items in one of the three formats appeared in each ELA form. Multiple forms were administered, each one to a different sample of students. An example of the different direction types for an ELA item and a mathematics item is presented in Exhibit 3.

Exhibit 3. Example of the Types of Instructions Under the Minimal, Concise, and Extensive Instruction Condition for the Item That Follows

**ELA Example**

**Minimal Directions**
Drag the **best** transition word to each blank in the paragraph.

**Concise Directions**
Complete the paragraph by selecting the **best** transition word that fits in each blank. Drag each transition word you selected to the correct blank in the paragraph.

**Extensive Directions**
There are six transition words in the text box. Complete the paragraph correctly by choosing a transition word that **best** fits each blank. Drag the transition word you selected from the text box to the correct blank in the paragraph.

|  |
|---|
|  |

It was winter. The cold wind was blowing and snow was covering the ground. Sarah gazed out the window and saw a bird trying to find food. She wanted to help the bird. After thinking for a while, Sarah decided to make a pinecone bird feeder. First, she tied a string to the top of a pinecone. _____, she covered the pinecone with peanut butter. After this, she placed the pinecone in the freezer. Later, she rolled the pinecone in birdseed. _____, she placed the pinecone bird feeder on a tree for the birds.

Mathematics Example

<u>**Minimal Directions**</u>

Drag numbers to make the equations true.

<u>**Concise Directions**</u>

Move numbers to make the equations true.

Drag the numbers to the answer space.

<u>**Extensive Directions**</u>

Drag numbers to make the equations true.

Each number can be used only once. To use a number, drag it to the appropriate box in an equation.

$$\sqrt{\boxed{\phantom{xxx}}} = \boxed{\phantom{xxx}}$$

$$\sqrt[3]{\boxed{\phantom{xxx}}} = \boxed{\phantom{xxx}}$$

## Results

Table 19 provides a count of the students in a grade band, by content area and direction type.

Table 19. Sample Sizes by Content Area, Direction Type, and Grade Band

| Content | Direction Type | Grade 3 | Grade 4 | Grade 6 | Grade 7 | Grade 11 |
|---------|----------------|---------|---------|---------|---------|----------|
| ELA | Minimal | 14 | 12 | 14 | 14 | 10 |
| ELA | Concise | 12 | 15 | 12 | 12 | 14 |
| ELA | Extensive | 18 | 17 | 15 | 7 | 6 |
| Mathematics | Minimal | 4 | 11 | 8 | - | 18 |
| Mathematics | Concise | 20 | 4 | 27 | - | 27 |
| Mathematics | Extensive | 19 | 4 | 27 | - | 16 |

Table 20a shows the percentage of students receiving full credit for the ELA items by direction type, item type, and grade band. In grade band 3, "select text" items were more challenging than "reorder text" items. This was especially true when the directions were "concise." With the "reorder text" items the grade band 3 students did less well with minimal directions. The grade band 11 students also had some difficulty with the "reorder text" items when the directions were "extensive." For the other grade bands, neither the level of instruction nor the item type showed a differential effect.

Table 20a. Percentage of Students Who Received Full Credit on ELA Items by Direction Type and Grade Band

| ELA | | Grade Band | | | | |
|-----|--|---|---|---|---|----|
| Direction Type | Item Type | 3 | 4 | 6 | 7 | 11 |
| Minimal | Reorder Text | 40 | | | | 71 |
| Concise | Reorder Text | 100 | | | | 59 |
| Extensive | Reorder Text | 67 | | | | 33 |
| Minimal | Select and Order | | 69 | | | |
| Concise | Select and Order | | 75 | | | |
| Extensive | Select and Order | | 53 | | | |
| Minimal | Select Text | 33 | | 100 | 41 | |
| Concise | Select Text | 0 | | 100 | 60 | |
| Extensive | Select Text | 38 | | 100 | 50 | |

In mathematics, a low percentage of students received full credit for "placing points" under the minimal and concise directions in grade band 11 (Table 20b). However, under extensive directions all students received full credit. With "placing points and tiling" items a higher percentage of students received full credit as the amount of instructions were reduced (grade band 6). "Select and order" items were difficult (grade bands 6 and 11) regardless of the direction type; however, no direction type proved better than another. The "select defined partition" items and the "straight lines" items showed high percentages of students receiving the maximum score, but the direction type did not make a difference. "Vertex-based quadrilateral" items seemed to benefit from minimal directions in grade band 11. Finally, "tiling" items were generally difficult, but no benefit was shown for different types of directions. The incompleteness of the data limits other comparisons.

Table 20b. Percentage of Students Who Received Full Credit on Different Types of Mathematics Items, by Direction Type and Grade Band

| Direction | Template | Grade Band | | | |
|---|---|---|---|---|---|
| | | 3 | 4 | 6 | 11 |
| Minimal | Placing Points | | | | 21 |
| Concise | Placing Points | | | | 21 |
| Extensive | Placing Points | | | | 100 |
| Minimal | Placing Points and Tiling | | | 67 | |
| Concise | Placing Points and Tiling | | | 57 | |
| Extensive | Placing Points and Tiling | | | 38 | |
| Minimal | Select and Order | | | | 44 |
| Concise | Select and Order | | | 32 | 43 |
| Extensive | Select and Order | | | 33 | 0 |
| Minimal | Select Defined Partitions | 100 | 70 | | |
| Concise | Select Defined Partitions | 76 | 100 | | |
| Extensive | Select Defined Partitions | 71 | 83 | | |
| Extensive | Single Ray | | | 15 | |
| Minimal | Straight Lines | | 100 | | 100 |
| Concise | Straight Lines | 100 | 100 | | 100 |
| Extensive | Straight Lines | 100 | | | |
| Extensive | Straight Line and Tiling | | | 29 | |
| Concise | Tiling | 19 | | | |
| Extensive | Tiling | 20 | 20 | | |
| Minimal | Vertex-Based Quadrilaterals | | | | 69 |
| Concise | Vertex-Based Quadrilaterals | | | 64 | 88 |
| Extensive | Vertex-Based Quadrilaterals | 30 | | | |

*Understanding instructions*

In ELA (Table 21a), for most item type/direction type/grade band combinations few students had difficulty understanding instructions. Cases in which difficulties were mentioned included about 50 percent of the students in grade band 4 with both minimal and extensive instructions for the "select and order" items. This was also true in grade band 3 for the "reorder text" items with extensive instructions and for the "select test" items with concise and extensive instructions. Finally, in grade band 11 the "reorder text" items with minimal and concise instructions elicited more comments.

In mathematics (Table 21b), the cases in which more comments were made about the instructions included "placing points" with minimal and concise instructions (grade band 11), "single ray" items with extensive instructions (grade band 6), "straight lines" items with extensive instructions, and "vertex-based quadrilateral" items with extensive instructions (grade band 3). The single ray item with extensive instructions in grade band 6 stood out as an item in which instructions were not well understood. ("Weren't totally sure how instructions were to be completed.") The percentage of students getting the maximum score on this item type was also low.

Table 21a . Percentage of Students Who Express the Difficulties in Understanding Each Type of Instruction for Each TE Type in Their Think-Alouds (ELA)

| ELA | | Grade Band | | | | | | | | | |
| | | 3 | | 4 | | 6 | | 7 | | 11 | |
| Direction Type | Item Type | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Minimal | Reorder Text | 0 | 0 | | | | | | | 0 | 20 |
| Concise | Reorder Text | | 25 | | | | | | | 0 | 20 |
| Extensive | Reorder Text | 33 | 12 | | | | | | | 0 | 0 |
| Minimal | Select and Order | | | 50 | 11 | | | | | | |
| Concise | Select and Order | | | 0 | 6 | | | | | | |
| Extensive | Select and Order | | | 44 | 0 | | | | | | |
| Minimal | Select Text | 0 | 0 | | | | 0 | 6 | 0 | | |
| Concise | Select Text | 33 | | | | | 14 | 0 | 22 | | |
| Extensive | Select Text | 25 | 40 | | | | 19 | 10 | 10 | | |

Table 21b. Percentage of Students Who Express the Difficulties in Understanding Each Type of Instructions for Each TE Type in Their Think-Alouds (Mathematics)

| Math | | Grade Band | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | 3 | | 4 | | 6 | | 11 | |
| Direction Type | TE type | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit | Non-Full Credit | Full Credit |
| Minimal | Placing Points | | | | | | | 55 | 33 |
| Concise | Placing Points | | | | | | | 67 | 0 |
| Extensive | Placing Points | | | | | | | | 20 |
| Minimal | Placing Points and Tiling | | | | | 0 | 0 | | |
| Concise | Placing Points and Tiling | | | | | 33 | 25 | | |
| Extensive | Placing Points and Tiling | | | | | 7 | 33 | | |
| Minimal | Select and Order | | | | | | | 14 | 9 |
| Concise | Select and Order | | | | | 13 | 8 | 4 | 10 |
| Extensive | Select and Order | | | | | 12 | 8 | 12 | |
| Minimal | Select Defined Partitions | | 0 | 0 | 0 | | | | |
| Concise | Select Defined Partitions | 14 | 9 | | 0 | | | | |
| Extensive | Select Defined Partitions | 25 | 13 | 0 | 20 | | | | |
| Extensive | Single Ray | | | | | 82 | 33 | | |
| Minimal | Straight Lines | | | | 25 | | | | |
| Concise | Straight Lines | | | | 0 | | | 100 | 0 |
| Extensive | Straight Lines | | 50 | | | | | | |
| Extensive | Straight Lines and Tiling | | | | | 0 | 0 | | |
| Concise | Tiling | 15 | 0 | | | | | | |
| Extensive | Tiling | 6 | 0 | 0 | 0 | | | | |
| Minimal | Vertex-Based Quadrilaterals | | | | | | | 25 | 0 |
| Concise | Vertex-Based Quadrilaterals | 30 | 0 | | | 67 | 12 | 33 | 14 |
| Extensive | Vertex-Based Quadrilaterals | 43 | 17 | | | | | | |

*Difficulty Using the Computer*

The results for ELA related to difficulty using the computer were mixed (Table 22). In grade band 3 under minimal directions for both "select text" and "reorder text" items, the students seemed to have difficulty using the computer. The grade band 11 students seemed to have some difficulty with the "reorder text" items.

Table 22. Percentage of Students Who Said They Had Trouble Using the Computer (ELA)

| Direction Type | Item | Grade | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Characteristic | 3 | 4 | 6 | 7 | 11 |
| Minimal | Select Text | 43 | | 4 | 0 | |
| Concise | Select Text | 25 | | 0 | 4 | |
| Extensive | Select Text | 19 | | 8 | 0 | |
| Minimal | Select and Order | | 22 | | | |
| Concise | Select and Order | | 25 | | | |
| Extensive | Select and Order | | 16 | | | |
| Minimal | Reorder Text | 31 | | | | 25 |
| Concise | Reorder Text | 11 | | | | 48 |
| Extensive | Reorder Text | 24 | | | | 30 |

Most students in mathematics had little trouble using the computer with mathematics items.

## Summary

In most cases in ELA the level of instruction did not have an influence. For most grade bands and item types, neither the level of instruction nor the item type had a differential effect in ELA. Cases in which differences were observed included "select text" items when the directions were "concise" (grade band 3). With the reorder text items the grade band 3 students did less well with minimal directions. The grade band 11 students also have some difficulty with the "reorder text" items when the directions were "extensive."

In mathematics, the level of instruction also did not make a difference for many of the item types and grade bands. "Select and order" items were difficult (grade bands 6 and 11) regardless of the direction type; however, no direction type proved better than another. High percentages of students received full credit on the "select defined partition" items and the "straight lines" items; however, the direction type did not make a difference. Finally, "tiling" items were generally difficult, but no benefit was shown for different types of directions. Places where differences were observed included "placing points" under the minimal and concise directions in grade band 11; however, under extensive directions all students received the maximum score. In working with "placing points and tiling" items, a higher percentage of students received full credit with fewer instructions (grade band 6). Finally, "vertex-based quadrilateral" items seemed to benefit from minimal directions in grade band 11.

The results for ELA related to trouble using the computer were mixed. In grade band 3 under minimal directions with both select text and reorder text items the students seemed to have difficulty using the computer.  The grade band 11 students seemed to have some difficulty with the "reorder text" items. Mathematics students did not seem to have any problems using the computer.

### ELA Questions, Passage Processing

*Research Question 7: Smarter currently intends to administer the passage first, and then administer the items one item at a time. Does this affect student performance?*

Smarter Balanced is interested in the possibility of administering items adaptively within a passage. This would require administering items sequentially so that the ability estimate could be updated after each item. Presenting items one at a time may take longer, and students may object to not knowing what is coming next. This question is designed to assess whether administering an item set takes longer when the items are presented sequentially and whether there is a difference in confusion or frustration level when students are presented a passage and all the items together or are presented a passage with the items then being presented one at a time. The item sets were not administered adaptively.

Two sets of items were created for a given test form. Both sets contained passages of equivalent length and difficulty as well as items of equivalent difficulty.[2] The first set in a form presented the passage with all the items together. The second set presented the passage with the items presented one at a time.

The forms were administered, within grade band, to different samples of students. Each sample contained both a general education group (Gen Ed) and a group that received ELL students. One sample was timed without thinking aloud during the administration. Each item set in these forms was separately timed. This sample provided timing information only. The second sample involved thinking aloud while responding to the questions and was not timed. Forms were constructed in ELA at three grade bands: grades 3–5 (referred to as grade band 3), grades 6–8 (referred to as grade band 6), and grades 10 and 11 (referred to as grade band 11).

The primary questions of interest were:

1. Does presenting the items individually after the passage appear to take longer (timed condition)?
2. Does presenting the items individually after the passage increase the student's negative emotional states (e.g., frustration, confusion; think-aloud condition)?
3. Do students prefer one approach or another (think-aloud condition)?

---

[2] Comparable passage difficulty was achieved through the use of readability and lexile measures. Comparable item difficulty was achieved through depth of knowledge (DOK) measures.

## Results

Table 23 shows the sample sizes taking each form of the tests, by grade band, for the ELL and Gen Ed samples. Sample sizes are smaller for the ELL sample in grade band 11.

Table 23. Student Counts by Grade Band, Testing Population, and Testing Condition

|  |  | Grade Band | | |
|---|---|---|---|---|
|  |  | 3 | 6 | 11 |
| Timed | Gen Ed | 9 | 6 | 8 |
|  | ELL | 8 | 4 | 1 |
| Think-Aloud | Gen Ed | 6 | 6 | 7 |
|  | ELL | 8 | 7 | 2 |

Table 24 shows the time (in seconds) it took to complete the item sets when all items were presented together or items were presented one at a time, by grade band and sample. For the grade band 3 and grade band 11 samples, timing differed little whether the items were presented in one block or one at a time. However, for grade band 6, presenting the items one at a time took substantially longer. While there is some variability between the ELL and the Gen Ed samples, the differences are not large and show a similar pattern. Note that the grade band 11 ELL sample was a single student and is not presented to avoid misleading results.

Table 24. Average Time to Complete the Passage and Items, by Administration Format, Grade Band, and Sample

| Grade Band | Sample | N | Passage + All Items | Passage + One Item at a Time | Difference (All—One at a Time) |
|---|---|---|---|---|---|
| 3 | Gen Ed | 9 | 250 | 239 | 11 |
| 3 | ELL | 8 | 263 | 239 | 24 |
| 6 | Gen Ed | 6 | 401 | 462 | –61 |
| 6 | ELL | 5 | 336 | 465 | –129 |
| 11 | Gen Ed | 8 | 270 | 285 | –15 |

Tables 25 and 26 show whether the ELL or Gen Ed sample students expressed confusion (Table 25) or frustration (Table 26) with the passages or items. There appears to be slightly more confusion for both the Gen Ed and the ELL sample students in grade band 3 when all the items are presented together. However, similar frustration levels were observed under the two formats for the grade band 3 students. The grade band 6 ELL sample, showed similar patterns of frustration and confusion for the two presentation formats. However, the Gen Ed grade band 6 students showed slightly more confusion when the items were presented one at a time. The grade band 11 Gen Ed students showed similar levels of confusion and frustration under both administrative formats. The grade band 11 ELL sample included only two students and is not reported.

Table 25. Percentage of Students Expressing Confusion with the Different Components of the Test by Administration Format, Grade Band, and Sample

| Sample | Test Component | All Items | | | One at a Time | | |
|---|---|---|---|---|---|---|---|
| | | Grade Band | | | Grade Band | | |
| | | 3 | 6 | 11 | 3 | 6 | 11 |
| Gen Ed | Passage | 33 | 29 | 17 | 0 | 43 | 14 |
| | Items | 25 | 30 | 17 | 9 | 36 | 18 |
| ELL | Passage | 50 | 50 | | 25 | 50 | |
| | Items | 32 | 50 | | 16 | 50 | |

Table 26. Percentage of the Students Expressing Frustration with the Different Components of the Test, by Administration Format, Grade Band, and Sample

| Sample | Test Component | All Items | | | One at a Time | | |
|---|---|---|---|---|---|---|---|
| | | Grade Band | | | Grade Band | | |
| | | 3 | 6 | 11 | 3 | 6 | 11 |
| Gen Ed | Passage | 0 | 29 | 17 | 0 | 29 | 14 |
| | Items | 13 | 18 | 17 | 13 | 11 | 14 |
| ELL | Passage | 13 | 38 | | 13 | 38 | |
| | Items | 3 | 41 | | 3 | 50 | |

Table 27 presents the average score students obtained for the think-aloud protocols. The grade band 6 students tended to score higher when the items were presented all at one time (for both the Gen Ed students and the ELL students). The grade band 3 students scored higher when the items were presented one at a time, regardless of sample or testing condition. The grade band 11, Gen Ed students scored higher when the items were presented one at a time, while the grade band 11, ELL sample students scored higher when the items were presented all at one time, though the latter sample size is small.

Table 27. Average Score, by Administration Format, Grade Band, and Sample

| | All Items at Once | | | One Item at a Time | | |
|---|---|---|---|---|---|---|
| | Grade Band | | | Grade Band | | |
| | | | | | | |
| Gen Ed | 2.2 | 3.0 | 1.8 | 2.5 | 2.3 | 2.5 |
| ELL | 2.4 | 2.9 | 2.0 | 2.5 | 1.7 | 1.5 |

Table 28 shows the preference for a presentation format. Both the ELL and Gen Ed grade band 3 students preferred to have the items presented one at a time. ("I preferred one at a time—less confusing than seeing too many questions," "One at a time made me less nervous about how many more there were," "I liked one at a time because it did not seem overwhelming.") Grade band 11 students (Gen Ed and ELL) had a slight bias toward having the items presented one at a time ("Let's me focus on that one question"). Conversely, grade band 6 Gen Ed students preferred to have the items presented together ("I liked them altogether," "This way I know I was on the same passage," "All together, you can refer to the questions while you read the passage," "I liked everything on one page because it was more easy," "With all together, I was able to refer back and I could see where I was going," "I liked altogether, though it was more confusing and distracting.") The grade band 6 ELL students were equally divided between the two formats.

Table 28. We Presented the Questions to You in Two Different Ways. Which Way Did You Prefer: All Together or One at a Time (Percent Responding)?

| | Grade Band | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 | | | 6 | | | 11 | | |
| Sample | All Together | No Preference | One at a Time | All together | No Preference | One at a Time | All Together | No Preference | One at a Time |
| Gen Ed | 33 | | 67 | 57 | 29 | 14 | 29 | 14 | 57 |
| ELL | 14 | | 86 | 43 | 14 | 43 | | 50 | 50 |

## Summary

We were interested in assessing whether there is a difference in timing and increased negative emotional states (confusion, frustration) when students are presented a passage with all the items or are presented a passage with the items presented one at a time. Forms were administered to two groups of students: a group that received English language accommodations and a Gen Ed group.

The time it took to complete the sets when all items were presented together or one at a time varied by grade band and sample. For the grade band 3 and grade band 11 samples, timing differed little whether the items are presented in one block or one at a time. However, for grade band 6, presenting the items one at a time took substantially longer for both the Gen Ed and ELL samples. While there is some variability between the ELL and the Gen Ed samples, the differences are not large and show the same pattern within grade band.

There appeared to be slightly more *confusion* for both the Gen Ed and the ELL samples in grade band 3 when all the items were presented together. However, similar *frustration* levels were observed under the two formats for the grade band 3 students. The grade band 6 ELL sample students showed similar patterns of frustration and confusion for the two presentation formats. However, the Gen Ed grade band 6 students showed slightly more confusion when the items were presented one at a time.

The grade band 6 students tended to score higher when the items were presented all at one time (for both the Gen Ed students and the ELL students). The grade band 3 students showed similar results, regardless of sample or administration format. The grade band 11, Gen Ed students scored higher when the items were presented one at a time, while the grade band 11 ELL sample students scored higher when the items were presented altogether.

Both the ELL and Gen Ed grade band 3 students preferred to have the items presented one at a time. Grade band 11 students had a slight bias toward having the items presented one at a time. Conversely, grade band 6 students preferred to have the items presented together.

*Research Question 8: Smarter intends to present relatively long passages. Do longer passages reduce student engagement?*

Smarter Balanced is interested in using passages that are longer than those presently used. The Smarter Balanced recommended passage lengths are: for grades 3–5: 450–562 words for short passages and 563–750 words for long passages; for grades 6–8: 650–712 words for short passages and 713–950 words for long passages; and for high school, 800–825 words for short passages and 826–1100 words for long passages. There is concern that the longer passages may tax the processing abilities of ELL and SWD students.

This question is designed to assess whether longer passages reduce student engagement, hamper the completion of the longer passages, or affect the depth of processing of the passage. Two sets of items were created. Both sets contained passages of equivalent difficulty with four items of equivalent difficulty attached to each passage. Both sets present the passage and all the items together. Each form contained a standard-length and an extended-length passage. The first set contained a passage of standard length. The second set contained a passage that is longer than

standard length (extended-length, the length equivalent to that intended for use by Smarter Balanced).

Forms were constructed in ELA at three grade bands: grade band 3–5 (referred to as grade band 3), grade band 6–8 (referred to as grade band 6), and grade band 10 and 11 (referred to as grade band 11). The design was intended to compare the performance of two groups of students—ELL/SWD and Gen Ed students—across three grade bands (3, 6, and 11). Thirteen students took the forms. Of these, nine were grade band 3 Gen Ed students. One grade band 3 student was classified ELL/SWD. The single grade band 6 student was an ELL/SWD student. The two grade band 11 students were Gen Ed students.

## Results

Table 29 shows the percentage of students whose engagement was improved or unaffected by the longer passage, by subgroup. All the ELL/SWD students were unaffected by the use of the longer passage. Gen Ed students did appear to be affected by the longer passage in grade bands 3 and 11. All the ELL/SWD students were able to read the entire passage regardless of passage length. Only about 25 percent of the grade band 3 Gen Ed students and none of the grade band 11 Gen Ed students were unaffected by the use of the longer passage (see Table 29; "I have to read the whole passage?"). The ELL/SWD students all demonstrated that the longer passage was processed at a deep level ("It was a good story"). However, only 43 percent of the Grade band 3, Gen Ed, students demonstrated a level of deep processing ("I learned many new things") and only 50 percent of the grade band 11 Gen Ed students demonstrated a level of deep processing (Table 31). The ELL/SWD students were not bored or distracted while reading either passage; however, some percentage of the Gen Ed students were bored regardless of the length of the passage.

Table 29. Percentage of Students Whose Engagement Is Improved or not Affected by the Longer Passage

| Subgroup | Grade Band | | |
|---|---|---|---|
| | 3 | 6 | 11 |
| GE | 25 | | 0 |
| ELL/SWD | 100 | 100 | |

Table 30. Percentage of Students Who Appear to Read the Entire Passage

| Standard Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 88 | | 100 |
| ELL/SWD | 100 | 100 | |

| Extended Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 88 | | 50 |
| ELL/SWD | 100 | 100 | |

Table 31. Percentage of Students Whose Think-Aloud Demonstrate Deep Processing as Assessed by the Interviewer

| Standard Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 43 | | 100 |
| ELL/SWD | 100 | 100 | |

| Extended Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 43 | | 50 |
| ELL/SWD | 100 | 100 | |

Table 32. Percentage of Students Who do not Appear Bored or Distracted

| Standard Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 63 | | 100 |
| ELL/SWD | 100 | 100 | |

| Extended Length | Grade Band | | |
|---|---|---|---|
| Subgroup | 3 | 6 | 11 |
| GE | 88 | | 50 |
| ELL/SWD | 100 | 100 | |

## Summary

Smarter Balanced is interested in using passages that are longer than those presently used. There is concern that the longer passages may tax the processing abilities of ELL and SWD) students. This question is designed to assess whether longer passages reduce student engagement, hamper the completion of the longer passages, or affect the depth of processing of the passage. The design was intended to compare the performance of two groups of students—ELL/SWD and Gen Ed students— across three grade bands (3, 6, and 11). Two sets of items were created. Both sets contained passages of equivalent difficulty with four items of equivalent difficulty attached to each passage. Both sets present the passage and all the items together. Both the standard-length and the extended-length passage were included in a given form and administered to the same student.

All the ELL/SWD students were unaffected by the use of the longer passage. They were able to read the entire passage regardless of passage length and demonstrated that the longer passage was processed at a deep level. The ELL/SWD students also were not bored or distracted while reading either passage.

On the contrary, Gen Ed students did appear to be affected by the longer passage in grade bands 3 and 11. About 75 percent of the grade band 3 students and all of the grade band 11 students were affected by the use of the longer passage. Only 43 percent of the Grade band 3 Gen Ed students demonstrated a level of deep processing and only 50 percent of the grade band 11 Gen Ed students demonstrated a level of deep processing. Also, some percentage of the Gen Ed students were bored, regardless of the length of the passage

*Research Question 9: How long does it take for students to read through complex texts, performance tasks, etc.? Is timing affected by the way students are presented the passage and items?*

One way of making items more difficult is to increase their complexity. Complex items often take longer to solve or answer. In computer adaptive tests, added complexity may decrease the time a high ability student has to complete the test if the items are made more difficult through increased complexity. This potentially creates some fairness issues in an adaptive test if there is a time limit on the test. This question was designed to assess the time it takes for students to answer complex and simpler items. Complexity was defined as a function of the DOK demanded by the test question. It was hypothesized that more complex tasks would take more time.

Each ELA form had six items. These items varied in item complexity (simple or complex) and item format (SR, TE, or CR). The TE items were all "hot text" items. These items require the student to either highlight the text or drag the text to answer the item.

Forms were constructed in ELA at two grade bands: grade band 3–5 (referred to as grade band 3) and grade band 6 and 7 (referred to as grade band 6). Two forms were administered in grade band 3. One form was administered in grade band 6.

## Results

Eight students took the grade band 3 forms with four students taking each form, and two students took the grade band 6 form.

Table 33 presents the average time (in seconds) a student took to answer an item. SR items were answered in the shortest time. HT items took about one minute longer than the SR items. CR items took the most time to answer, about 75 seconds longer than the "hot text" items. With the exception of the complex CR item administered to grade band 6 students, item complexity did not seem to have an impact on item performance. (An interviewer commented, "Student took about the same time for complex and easy items.")

Table 33. Average Time (in seconds) to Answer an Item by Grade Band, Item Type, and Item Complexity

| | | | Grade Band | |
|---|---|---|---|---|
| | | | 3 | 6 |
| Item Format | Difficulty | Item | Avg. Time | Avg. Time |
| SR | Simple | 1 | 49 | 52 |
| SR | Complex | 2 | 29 | 59 |
| TE (HT) | Simple | 3 | 83 | 126 |
| TE (HT) | Complex | 4 | 96 | 123 |
| CR | Simple | 5 | 182 | 168 |
| CR | Complex | 6 | 158 | 185 |

Table 34 presents a summary of the average time students took to complete complex and simple items across item types by grade band. Complex items seemed to have more impact in grade band 6, but there is no evidence that complex items, as defined here, take longer than simpler items.

Table 34. Interviewer's Summary of Item Timing by Grade Band and Item Difficulty

| | Grade Band | |
|---|---|---|
| | 3 | 6 |
| Difficulty | Avg. Time | Avg. Time |
| Simple | 104 | 115 |
| Complex | 94 | 126 |

## Summary

It was hypothesized that more complex items would take longer to complete than simpler items. No evidence was found to support this hypothesis. In terms of the time spent on an item, SR items were answered in the shortest time. "Hot text" items took about one minute longer than SR items. CR items took the most time to answer, about 75 seconds longer then the "hot text" items.

## Effective Communication of Mathematics

*Research Question 10: Working mathematics problems on computer: Communicating mathematics on computer—feasibility of measuring student understanding of items for Claims 2–4 on computer.*

With paper tests some students write in their test books while working out mathematics problems. When mathematics items are presented on computer, scratch paper is often provided if students want to transfer the problem to paper and work it out there. Because scratch paper is often destroyed after an online testing session, the degree to which scratch paper is used is not known; neither is the importance of scratch paper in working out a problem (or potentially for use in scoring). This research question examines the need for paper when solving mathematics problems. Forms were constructed at four grade bands: grade band 3 and 4 (referred to as grade band 3), grade band 6 and 7 (referred to as grade band 6), grade band 7 and 8 (referred to as grade band 7), and grade band 11 (referred to as grade band 11) to investigate whether the scratch paper usage was uniform or varied by educational level.

Each student was presented with three grade-appropriate items. The interviewer recorded whether the student made a comment, and the nature of the comment, while working the mathematics problems. The students first tried to work the problem without paper. Scratch paper was then offered to the student to rework the problem, if desired. The interviewer noted whether students chose to add anything additional and noted the nature of the addition (more text, equations, graphics). Note that there were only three comments for the third item in the lowest grade band, 3.

## Results

Twenty students were administered the grade band 3 form, 37 students were administered the grade band 6 form, 21 students were administered the grade band 7 form, and 19 students were administered the grade band 11 form.

Table 35 shows the percentage of comments made for an item and the type of comment made. Two types of comments were of interest: did the students who wanted paper draw a picture or write an equation or did they find the online system difficult to use. The lowest grade band students (grade band 3) did not need paper to solve any of the problems (Table 635. Some students in the highest grade band (grade band 11) commented that they would like to draw a picture for the items they were administered (15–30 percent). ("I wanted to graph the area.") There was also one item (Item 2) for which about 15 percent of students wanted paper to write equations. About 5–10 percent of students in each grade band found the online system difficult to use. ("Confused me, I didn't know how to write an equation," "Tried the keypad, but it wouldn't work," "It was much easier with paper.") The strongest result came from the grade band 6 and grade band 7 groups, where 30 to 42 percent of the sample, respectively, indicated that they wanted to write an equation. Between 3 and 23 percent of the grade band 6 and 7 groups also indicated that they wanted to draw a picture. This may be a function of newly introduced algebra concepts for this group.

Table 35. Percentage of Comments for an Item, by Question Type and Grade Band

| Question | Item | Grade Band | | | |
|----------|------|---|---|---|----|
| | | 3 | 6 | 7 | 11 |
| Picture | 1 | 5 | 0 | 23 | 32 |
| | 2 | 15 | 12 | 3 | 16 |
| | 3 | 0 | 4 | 6 | 16 |
| System Difficulty | 1 | 5 | 9 | 10 | 5 |
| | 2 | 11 | 3 | 10 | 5 |
| | 3 | 0 | 4 | 7 | 5 |
| Equation | 1 | 0 | 31 | 45 | 6 |
| | 2 | 0 | 32 | 34 | 16 |
| | 3 | 0 | 29 | 43 | 6 |

Table 36 shows the nature of the student comments made on paper and whether the additional information recorded on the paper improved the response according to the rubric. For all grade bands the additional information recorded on the paper included a graphic. In grade bands 6, 7, and 11, the additional information recorded on paper included an equation. The grade band 6, 7, and 11 groups provided additional information on paper that improved the response according to the rubric. For example, one administrator noted, "When given paper, she was able to do the proper equation and solve for x. She was more confident with paper and pencil." The number of cases in which improvement was observed varied by item. For grade band 6, item 2, about 11 percent of the responses were improved when scratch paper information was taken into account during scoring. For grade band 11, item 3, about 16 percent of the responses were improved when scratch paper information was taken into account during scoring. Responses to all items in grade band 7 were improved when scratch paper information was taken into account. The improvement for this group ranged between 10 and 20 percent across items.

Table 36. Percentage of Changes Made When Paper Was Introduced

| Nature of Students' Changes | Item | Grade Band | | | |
|---|---|---|---|---|---|
| | | 3 | 6 | 7 | 11 |
| No Additions Made | 1 | 80 | 57 | 71 | 53 |
| | 2 | 60 | 65 | 67 | 63 |
| | 3 | 10 | 32 | 52 | 58 |
| Addition Included Graphic | 1 | 5 | 3 | 33 | 37 |
| | 2 | 15 | 5 | | 11 |
| | 3 | | | 10 | 32 |
| Addition Included Equation | 1 | | 22 | 19 | 5 |
| | 2 | 20 | 16 | 38 | 16 |
| | 3 | | 19 | 38 | 11 |
| Addition Improved Response According to Rubric | 1 | | 11 | 14 | |
| | 2 | | 3 | 29 | |
| | 3 | | | 24 | 16 |

The interviewer's comments suggested that most students in grade band 3 (75 percent) and grade band 11 (63 percent) were able to accurately respond to the mathematics items they saw only using the online text editor. However, fewer than half of the students in grade band 6 (45 percent) could accurately respond to questions using only the text editor and only 13 percent of the students in grade band 7 were observed to be able to accurately respond to questions using only the text editor. One student commented, "It's much easier with paper."

## Summary

The general conclusion is that a subset of students benefit from being able to work mathematics problems on paper. It appears to be especially important when students are beginning to learn algebra concepts.

Grade band 3 students did not need paper to work the problems. However, in the grade band 6 and grade band 7 groups, 30–42 percent of students indicated that they wanted to write an equation. In grade bands 6, 7, and 11, the additional information recorded on paper would have improved the

response according to the rubric. Responses for specific items in grade bands 6 and 11 were improved by 15 percent of the students and responses for all items in grade band 7 were improved when information on the scratch paper was taken into account. Improvement for this group ranged between 10 and 20 percent of the responses. This was supported by interviewer observations. About 5–10 percent in each grade band found the online system difficult to use, but few specifics were recorded.

*Research Question 11: Usability of equation editor tool—can students use the tool the way it is meant to be used?*

Although students begin to use technology at a very early age, it is prudent to verify that young students are able to use the assessment interface to be used during testing. This question sought to evaluate the ability of grade 3–5 students to use the equation editor tool to be included in the Smarter Balanced delivery system. Three mathematics items were presented to the students (*N*=33). The first item only required the student to copy his or her response. The second item was a simple mathematics item and the third item was a more challenging mathematics item. The first item would demonstrate whether the student could use the equation editor tool. The second and third items would provide evidence of whether the ability to use the tool interacted with item difficulty.

### Results

Between 15 and 30 percent of the students indicated that they had difficulty using the equation editor. About 30 percent had trouble just copying the answer, as required by item 1. The examiners assessed that 35 percent had difficulty using the equation editor and that only 40–57 percent of the students would get a given item correct. Students had more difficulty with the more challenging items. A summary of representative comments made by students about the equation editor during the administration of the think-aloud protocol is presented below:

1. Clicked on the + sign, but it didn't work, twice.
2. How do I choose the numbers?
3. I needed paper to make a picture.
4. How do I use the number pad?
5. I tried to use the numbers on the keyboard, but wouldn't work.
6. Some symbols didn't respond to first click.
7. I had trouble getting bottom half of fraction to record.
8. Unclear what possible value meant.
9. I didn't see decimal point down there [due to scrolling].
10. Couldn't find x symbol.
11. Unclear whether to click and drag or type.
12. Would rather type than use a mouse.
13. Difficult to use fraction tool.

### Summary

Elementary students had some difficulty using the equation editor. Between 15 and 30 percent of the students indicated that they had difficulty using the equation editor. The examiner's assessment

concurred that about 35 percent had difficulty using the equation editor and that about 50 percent of the students would get a given item correct.

*Research Question 12: Intuitive understanding of the relationships in multiplying fractions.*

This question is designed to assess whether students with a strong understanding of fractions and the multiplication and division of fractions complete the items without performing the indicated multiplication. The task asked students to compare the size of a product to the size of one factor, on the basis of the size of the other factor, without performing the indicated multiplication. Also of interest was whether students who complete an item as intended (without using multiplication) spent less time on an item than those who did not. To investigate this question a single form was administered for grades 3–5.

### Results

The form was administered to 33 students at the elementary level. Table 37 compares those with a strong understanding of fractions with those who do not have a strong understanding of fractions and whether they completed the task with or without using multiplication. There does not appear to be a relationship between strength of understanding of fractions (multiplication and division) and whether they used multiplication to solve the problems.

Table 37. Strength of Understanding of Fractions and Whether Multiplication was Performed

| Item Number | Not Strong Understanding of Fractions | | Strong Understanding of Fractions | |
|---|---|---|---|---|
| | Performed Multiplication | Did not Perform Multiplication | Performed Multiplication | Did not Perform Multiplication |
| 1 | 9 | 7 | 8 | 1 |
| 2 | 9 | 8 | 9 | 1 |
| 3 | 10 | 6 | 6 | 1 |
| 4 | 6 | 7 | 10 | 1 |
| 5 | 7 | 7 | 9 | 1 |
| 6 | 4 | 6 | 15 | 0 |

Table 38 presents descriptive statistics for the timing of each item (in seconds). In addition to means, medians are reported because timing distributions tend to be highly skewed. On average, those who did not have to perform the multiplication completed the items in less time. The results for item 6 were comparable for the two groups.

Table 38. Comparison of the Time to Complete the Item for Those Who Did not Use Multiplication to Solve the Item and Those Who Did

| Item Number | Performed Multiplication | | | | Did not Perform Multiplication | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std Dev | Median | Range | Mean | Std Dev | Median | Range |
| 1 | 210 | 136 | 179 | 59–543 | 136 | 90 | 114 | 53–360 |
| 2 | 145 | 119 | 106 | 36–420 | 126 | 110 | 89 | 30–336 |
| 3 | 75 | 104 | 42 | 10–480 | 34 | 28 | 25 | 3–90 |
| 1 | 123 | 111 | 70 | 21–480 | 88 | 69 | 57 | 25–195 |
| 2 | 133 | 130 | 95 | 28–480 | 79 | 67 | 68 | 9–185 |
| 3 | 69 | 118 | 32 | 4–540 | 65 | 63 | 51 | 3–170 |

Table 39 shows the percentage of students answering the item correctly. The students tested generally found the items to be difficult. ("Multiplying fractions was hard.") Some students did not understand the inequality signs, while others did not understand improper fractions or how to make a whole number into a fraction. One interviewer commented that the "student had little or no understanding of fractions."

Table 39. Percentage of Students Answering an Item Correctly.

| Item Number | Percent |
|---|---|
| 1 | 17 |
| 2 | 20 |
| 3 | 28 |
| 4 | 42 |
| 5 | 26 |
| 6 | 33 |

About 69 percent of the students used multiplication to solve the problems (Table 40). Student comments support this. "I multiplied… each box and put them in the correct boxes (columns)." "I timesed [sic] the numbers." "I looked at each number expression and multiplied it in my head and moved it to where I thought it was right." "Some numbers on the bottom depends on the top number which is bigger or smaller." Only about 40 percent of the students understood fractions or at least the multiplication of fractions. The examiner's comments (Table 41) concur with this conclusion.

Table 40. Percentage of Students Using Multiplication to Solve the Items

| Item Number | Yes |
|:-----------:|:---:|
| 1 | 68 |
| 2 | 70 |
| 3 | 75 |
| 4 | 72 |
| 5 | 73 |
| 6 | 81 |

Table 41. Interviewer's Assessment of: (1) Whether the Student Used Multiplication and (2) Whether the Student Had a Strong Understanding of Fractions

| Summary | Percent |
|:-------:|:-------:|
| Did student use multiplication? | 72 |
| Did student have a strong understanding of fractions (multiplication/division)? | 40 |

## Summary

There seemed to be little relationship between whether a student has a strong understanding of the multiplication and division of fractions and whether he or she used multiplication to solve the items. However, students who did not have to perform the multiplication completed the items in less time than students who had to perform the multiplication. While most students said they understood the questions, 70 percent had to use multiplication to solve them. Only about 40 percent of the students had a firm understanding of the multiplication/division of fractions, according to the interviewers.

## Special Populations

*Research Question 13: Contextual glossaries are item-specific glossaries that provide a definition of a word that is targeted to, and appropriate for, the context in which the word is used in the item. Are these a fair and appropriate way to support students who need language support?*

This question addressed the efficacy of the use of contextual glossaries with non-native (Spanish) speakers (see Exhibit 4 for an example of a contextual glossary item) when solving mathematics problems. A contextual glossary item contains highlighted words when presented online. Clicking any of these highlighted items produces a list of all highlighted words in the item with Spanish definitions for each. Two sets of items were created that were parallel in difficulty. The first set of items contained no contextual glossaries with only single words translated. The second set of items

contained contextual glossaries. The interviewer was asked to determine whether the student was having trouble understanding a word and whether the contextual glossary aided in the interpretation of the word or sentence.

Only three ELL students participated: one from grade 3 and two from grade 6.

Exhibit 4. Example of a Contextual Glossary Item

1. A roller coaster has a large rise and drop followed by a complete circle. The following diagram shows measurements for the track. An extra 20 feet are needed for cutting and welding. How many feet of track should be ordered? (Use π = 3.14)

A.  280 feet

B.  407 feet

C. 415.6 feet

D. 1,537.4 feet

Glossary Window

**Roller coaster**

montaña rusa

**Rise**

subida

**Drop**

bajada
caída

**Complete**

completo
entero

**Diagram**

diagrama
quema
gráfico

**Track**

vía
riel

**Cutting**

cortar

**Welding**

cortar

## Results

The grade 3 student had trouble understanding a few items, but had few word confusions. For the second set of items, this student used the contextual glossaries for one item but not for the other items. The student said that there was not a problem understanding the items because the student used "sentence context" to answer them, or the words the student didn't know weren't in the glossary so the student stopped using it. In terms of scoring, this student answered two of the three "translated" items correctly, but did not answer any of the "contextual glossary" items correctly, so the results are difficult to interpret as to whether the use of contextual glossaries aided the students' performance.

The two grade 6 students (one ELA form and one Math form) both had difficulty with the "translated" items in the first set with six or more word confusions each for most items. Both students found the contextual glossary useful to some degree, though not for all items. ("The words I don't know aren't in the glossary.") However, the interviewers suggested that the use of the contextual glossary improved the performance for both grade 6 students. Though the ELA student got all questions incorrect, the interviewer believed that this was mainly due to careless mistakes and that the student used the glossary to help make sense of the key components of the questions and understood the procedures for answering the questions. The math student got two-thirds of the items correct when the items were translated, and one-third of the items correct when the contextual glossary was used. The student had difficulty understanding an essential word in one of the incorrect items. However, the interviewer commented that once he understood the words, he could confidently work on the problem and he knew how to proceed.

## Summary

In summary, contextual glossaries appeared to be somewhat effective when they were used, but the impact was not always reflected in the score the student received for an item. The contextual glossaries appeared to be incomplete in that they did not include words that the students needed. This limited the use of the glossaries in these situations. Interviewer's comments suggested that performance was improved when the students used the contextual glossaries.

*Research Question 14: Under what conditions do students with lower reading ability use text-to-speech (TTS) to help focus on content in ELA and mathematics? Is this affected by the quality of the voice-pack?*

TTS is a technology that can give students with low reading ability access to an assessment. For this technology to be effective the language produced from the voice-pack must be clear so that it can be understood. This is particularly true for non-native speakers of English.

This question is designed to assess whether students with lower reading ability and non-native speakers of English use TTS to help focus on content in ELA and mathematics. Only students familiar with TTS were included in the study. Overall, 77 students used TTS at least once. Among them, 58 students are LEP students, 13 students had reading difficulties (IEP), and six students were Gen Ed students.

Forms were constructed at three grade bands: grade band 3 (referred to as grade band 3), grade band 6 and 7 (referred to as grade band 6), and grade 11 (referred to as grade band 11). In ELA, four forms were administered with both high- and low-quality voice-packs. In mathematics, two forms were administered in grade bands 3 and 11. Only a single form was administered in grade band 6. For all mathematics forms only high-quality voice-packs were administered. In Tables 42–45, yellow shading denotes the use of high-quality voice-packs while a white background denotes the use of a low-quality voice-pack.

## Results

For ELA (Table 42), for all groups and grade bands, a high percentage of students tended to make comments indicating an improved focus on the content when the voice-pack was of high quality. About one-third of the students (except the Gen Ed grade band students) indicated that TTS kept their focus on content even when low-quality voice-packs were used. For ELA, students in all groups tended to make greater use of TTS when the voice-pack was of high quality.

About 50 percent of the LEP students in mathematics in grade bands 3 and 11 made comments indicating that TTS helped them focus on content. All of the LEP grade band 6 group and the IEP students in grade band 3 found that TTS helped them focus on content. ("It made me think about the question.") The Gen Ed students in grade band 3 found that TTS helped them focus on content; however, the Gen Ed grade band 6 students did not find TTS useful.

Table 42. Percentage of TTS Students Who Made Any Comment Indicating That He/She Is Mainly Focused on the Content of the Item, by Content, Voice-Pack Quality, Sample, and Grade Band

| Content | Voice Pack Quality | Grade Band | LEP | | | IEP | | | Gen Ed | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3 | 6 | 11 | 3 | 6 | 11 | 3 | 6 | 11 |
| ELA | Low | | | 32 | 39 | 35 | | | | 0 | |
| | High | | 36 | 67 | 100 | 100 | | | | | 100 |
| Mathematics | Low | | | | | | | | | | |
| | High | | 50 | 100 | 60 | 100 | | | 100 | 0 | |

Table 43 shows the percentage of students who answered the items correctly, averaged across items. In ELA, the grade band 6 and 11 LEP students and the grade band 3 IEP students found the items more difficult using a low-quality voice-pack. The Gen Ed grade band 6 ELA students were not administered a high quality voice-pack. In the LEP grade band 6 group, about half the students answered an item correctly using the high-quality voice-pack. The percentage answering an item correctly was close to 75 percent for the other LEP grade bands and the grade band 3 low-level reading students when the high-quality voice-pack was used.

In mathematics, in grade band 3, about 40 percent of the LEP students answered an item correctly. For the other grade bands, for the LEP and IEP samples, no items were answered correctly, even with the high-quality voice-packs. This was also true for the Gen Ed grade band 3 students. However, the general education students in grade band 6 answered all the items correctly.

Table 43. Percentage of TTS Students Who Answered the Items Correctly by Content, Voice-Pack Quality, Sample, and Grade Band

| Content | Voice Pack Quality | Grade Band | LEP | | | IEP | | | Gen Ed | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3 | 6 | 11 | 3 | 6 | 11 | 3 | 6 | 11 |
| ELA | Low | | | 14 | 0 | 50 | | | | 75 | |
| | High | | 77 | 50 | 80 | 75 | | | | | 0 |
| Mathematics | Low | | | | | | | | | | |
| | High | | 40 | 0 | 0 | 0 | | | 0 | 100 | |

Tables 44 and 45 summarize the interviewer's assessment for ELA and mathematics related to whether TTS improved access to the content or was a distraction. TTS improved access in ELA regardless of the quality of the voice-pack. Greater access was achieved when high-quality voice-packs were used in ELA except in grade band 11. This is probably an artifact of the very small sample size. The low-quality voice-pack appeared less effective at providing access and was distracting in ELA, where the high-quality voice-pack was not distracting at all. One student said, "[I] didn't like using TTS … the sound was robotic and would break my concentration."

In mathematics, TTS helped to improve access for some grade band 3 LEP students, but not for middle- and upper-level LEP students or the IEP or Gen Ed grade band 3 students. All the Gen Ed, IEP, and grade band 6 LEP students found the high-quality voice-pack distracting in mathematics. This was in part a function of trying to describe a table verbally. ("When TTS read the chart aloud, I got lost in the numbers and couldn't figure out what the question was asking.")

Table 44. Assessment by the Interviewer of the Percentage of TTS Students Whose Access to Content Was Improved by the Use of TTS by Content, Voice-Pack Quality, Sample, and Grade Band

| Content | Voice Pack Quality | Grade Band | LEP | | | IEP | | | Gen Ed | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3 | 6 | 11 | 3 | 6 | 11 | 3 | 6 | 11 |
| ELA | Low | | | 57 | 75 | 79 | | | | 100 | |
| | High | | 76 | 100 | 33 | 100 | | | | | 100 |
| Mathematics | Low | | | | | | | | | | |
| | High | | 43 | 0 | 0 | 0 | | | 0 | 0 | |

Table 45. Assessment by the Interviewer of the Percentage of TTS Students Who Were Distracted by TTS, by Content, Voice-Pack Quality, Sample, and Grade Band

| Content | Voice Pack Quality | Grade Band | LEP | | | IEP | | | Gen Ed | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3 | 6 | 11 | 3 | 6 | 11 | 3 | 6 | 11 |
| ELA | Low | | | 12 | 20 | 33 | | | | 0 | |
| | High | | 0 | 0 | 0 | 0 | | | | | 0 |
| Mathematics | Low | | | | | | | | | | |
| | High | | 44 | 100 | 40 | 0 | | | 100 | 100 | |

## Summary

TTS improved access in ELA regardless of the quality of the voice-pack. Greater access was achieved when high-quality voice-packs were used. LEP students and students with reading difficulties tended to benefit more from the use of TTS. Using TTS with high-quality voice-packs improved focus on content in ELA. The use of TTS with low-quality voice-packs tended to distract students in ELA, whereas high-quality voice-packs did not. In mathematics, access was improved only for grade band 3 students. All the Gen Ed, IEP, and grade band 6 LEP students found the high-quality voice-pack distracting. This was in part a function of trying to describe a table verbally.

## Final Summary

Smarter Balanced is moving toward an assessment model that is largely scored automatically and delivered adaptively on computer. The Smarter Balanced cognitive laboratories were conducted to investigate questions that arise from such an automated design. While think-aloud protocols are time consuming, they have the potential to provide a level of information not easily accessed through large-scale studies. However, the sample sizes are small. Therefore, should a more rigorous investigation of any of the research questions be of interest, specifically designed studies with large samples will be needed.

This report presents the results from 14 small think-aloud studies that addressed topics that pertain to an automated test delivery system.

1. Can non-constructed-response item formats assess components that have historically been believed to be measured only with CR items?
2. What is the optimal amount of direction to provide for TE items? Does this vary with grade level?
3. What is the appropriate degree of labeling to provide for MPSR items so that students know to complete all parts?
4. Does it matter whether items associated with a passage are presented in a single block or presented one item at a time? Are ELL students impacted by these different arrangements?
5. Do the longer passages favored by Smarter Balanced reduce student engagement?
6. How much time do items in different formats take to answer? Are ELL students affected more than general education students?
7. In mathematics, could information captured on scratch paper facilitate the working of a problem and benefit the performance and scoring of a student?
8. Do contextual glossaries help improve the performance of students with language disabilities?
9. Does TTS help focus students of low reading ability on the content of an item?
10. Can younger students effectively use the equation editor?
11. Mathematics intuition: Can students compare the size of a product to the size of one factor, on the basis of the other factor without multiplying?

On the whole, the cognitive laboratories were successful in providing answers to most of these questions. They provide a glimpse of issues that may exist and need to be investigated further. To investigate these issues more completely, larger-scale studies should be conducted.

**Appendix A**

Question 2. Full Claim Descriptions

| Content | Content Grade | Claim | Claim Description |
|---------|---------------|-------|-------------------|
| ELA | 3–5 | 1 | Students can read closely and analytically to comprehend a range of increasingly complex literary and informational text. |
| ELA | 3–5 | 2 | Students can produce effective writing for a range of purposes and audiences. |
| ELA | 3–5 | 3 | Students can employ effective speaking and listening skills for a range of purposes and audiences. |
| ELA | 3–5 | 4 | Students can engage in research/ inquiry to investigate topics and to analyze, integrate, and present information. |
| ELA | 6–8 | 1 | Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts. |
| ELA | 6–8 | 2 | Students can produce effective writing for a range of purposes and audiences. |
| ELA | 6–8 | 3 | Students can employ effective speaking and listening skills for a range of purposes and audiences. |
| ELA | 6–8 | 4 | Students can engage in research/ inquiry to investigate topics and to analyze, integrate, and present information. |
| ELA | 9–12 | 1 | Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts. |
| ELA | 9–12 | 2 | Students can produce effective and well-grounded writing for a range of purposes and audiences. |
| ELA | 9–12 | 3 | Students can employ effective speaking and listening skills for a range of purposes and audiences. |
| ELA | 9–12 | 4 | Students can engage in research/inquiry to investigate topics, and to analyze, integrate, and present information. |
| Math | 3–5 | 1 | Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency. |

| Content | Content Grade | Claim | Claim Description |
|---------|---------------|-------|-------------------|
| Math | 3–5 | 2 | Students can solve a range of well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. |
| Math | 3–5 | 3 | Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. |
| Math | 3–5 | 4 | Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems. |
| Math | 6–8 | 1 | Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency. |
| Math | 6–8 | 2 | Students can solve a range of well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. |
| Math | 6–8 | 3 | Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. |
| Math | 6–8 | 4 | Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems. |
| Math | 9–12 | 1 | Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency. |
| Math | 9–12 | 2 | Students can solve a range of well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. |
| Math | 9–12 | 3 | Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. |
| Math | 9–12 | 4 | Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems. |

Question 2. Full Target Descriptions

| Content | Grade Content | Grade Band | DOK | Claim | Target | Target Description |
|---------|---------------|------------|-----|-------|--------|--------------------|
| ELA | 3 | 3 | 2 | 2 | 3 | Write or revise one or more informational/explanatory paragraphs demonstrating ability to organize ideas by stating a focus, including appropriate transitional strategies for coherence, or supporting details, or an appropriate conclusion. |
| ELA | 6 | 6 | 2 | 1 | 11 | Use supporting evidence to justify interpretations or analyses of information presented or how information is integrated within a text (point of view; interactions among events, concepts, people, or ideas; author's reasoning and evidence). |
| ELA | 6 | 6 | 3 | 1 | 11 | Use supporting evidence to justify interpretations or analyses of information presented or how information is integrated within a text (point of view; interactions among events, concepts, people, or ideas; author's reasoning and evidence). |
| ELA | 7 | 6 | 2 | 2 | 3 | Apply a variety of strategies when writing or revising one or more paragraphs of informational/explanatory text organizing ideas by stating and maintaining a focus/tone, providing appropriate transitional strategies for coherence, developing a topic including relevant supporting evidence/vocabulary and elaboration, or providing a conclusion appropriate to purpose and audience. |
| ELA | 7 | 7 | 2 | 1 | 1 | Identify explicit textual evidence to support inferences made or conclusions drawn. |
| ELA | 8 | 7 | 2 | 1 | 1 | Identify explicit textual evidence to support inferences made or conclusions drawn. |
| ELA | 8 | 7 | 2 | 2 | 6 | Apply a variety of strategies when writing or revising one or more paragraphs of informational/explanatory text organizing ideas by stating and maintaining a focus/tone, providing appropriate transitional strategies for coherence, developing a topic including relevant supporting evidence/vocabulary and elaboration, or providing a conclusion appropriate to purpose and audience. |
| ELA | 11 | 11 | 2 | 1 | 1 | Cite explicit textual evidence to support inferences made or conclusions drawn about texts. |

| Content | Grade Content | Grade Band | DOK | Claim | Target | Target Description |
|---------|---------------|------------|-----|-------|--------|--------------------|
| ELA | 11 | 11 | 2 | 1 | 7 | Determine or analyze the figurative (e.g., euphemism, oxymoron, hyperbole, paradox), or connotative meanings of words and phrases used in context and the impact of those word choices on meaning and tone. |
| MATH | 3 | 3 | 2 | 1 | F | Develop understanding of fractions as numbers. |
| MATH | 4 | 3 | 2 | 1 | F | Extend understanding of fraction equivalence and ordering. |
| MATH | 4 | 4 | 2 | 1 | L | Draw and identify lines and angles, and classify shapes by properties of their lines and angles. |
| MATH | 4 | 4 | 3 | 4 | A | Apply mathematics to solve problems arising in everyday life, society, and the workplace. |
| MATH | 6 | 6 | 2 | 1 | E | Apply and extend previous understandings of arithmetic to algebraic expressions. |
| MATH | 6 | 6 | 3 | 2 | A | Apply mathematics to solve well-posed problems arising in everyday life, society, and the workplace. |
| MATH | 11 | 11 | 2 | 1 | I | Solve equations and inequalities in one variable. |
| MATH | 11 | 11 | 2 | 2 | A | Apply mathematics to solve well-posed problems arising in everyday life, society, and the workplace. |
| MATH | 11 | 11 | 3 | 2 | A | Apply mathematics to solve well-posed problems arising in everyday life, society, and the workplace. |

## Appendix B

## Demographic Information for Cognitive Laboratories

**Total Number of Students**: 774

**By Cognitive Lab Location:**
San Francisco, California: 80 (10%)
Monterey, California: 167 (22%)
Waterbury, Connecticut: 45 (6%)
Hartford, Connecticut: 26 (3%)
Pocatello, Idaho: 64 (8%)
District of Columbia: 31 (4%)
Honolulu, Hawaii: 43 (6%)
East Lansing, Michigan: 63 (8%)
Madison Heights, Michigan: 33 (4%)
Marquette, Michigan: 30 (4%)
Des Moines, Iowa: 52 (7%)
Pittsburgh, Pennsylvania: 76 (10%)
Columbia, South Carolina: 50 (6%)
Portland, Oregon: 14 (2%)

**By School Location:**
California: 243 (31%)
Connecticut: 71 (9%)
District of Columbia: 14 (2%)
Hawaii: 43 (6%)
Idaho: 64 (8%)
Iowa: 52 (7%)
Maryland: 12 (2%)
Michigan: 126 (16%)
Nevada: 4 (<1%)
Oregon: 12 (2%)
Pennsylvania: 76 (10%)
South Carolina: 50 (6%)
Virginia: 5 (<1%)
Washington: 2 (<1%)

**By Grade:**
Grade 3: 113 (15%)
Grade 4: 100 (13%)
Grade 5: 79 (10%)
Grade 6: 98 (13%)
Grade 7: 113 (15%)
Grade 8: 62 (8%)

Grade 9: 87 (11%)
Grade 10: 70 (9%)
Grade 11: 44 (6%)
Grade 12: 8 (1%)

**By Gender:**
Male: 393 (51%)
Female: 381 (49%)

**Language(s) Spoken at Home:**
English: 670 (87%)
Spanish: 100 (13%)
Chinese: 46 (6%)
Chaldean: 21 (3%)
Arabic: 18 (2%)
Albanian: 15 (2%)
Tagalog: 10 (1%)
German: 5 (<1%)
Vietnamese: 5 (<1%)
Hindi: 4 (<1%)
Korean: 4 (<1%)
Japanese: 3 (<1%)
Samoan: 3 (<1%)
Bengali: 2 (<1%)
Greek: 2 (<1%)
Ilocano: 2 (<1%)
Telegu: 2 (<1%)
Other: 14 (2%)
*Total percentage is more than 100% because more than one response could be selected.

**Language(s) Most Frequently Spoken:**
English: 707 (91%)
Arabic: 22 (3%)
Chinese: 18 (2%)
Chaldean: 16 (2%)
Spanish: 13 (2%)
Albanian: 3 (<1%)
Greek: 2 (<1%)
Tagalog: 2 (<1%)
Other: 7 (1%)
*Total percentage is slightly over 100% because some parents added an additional language in the comment section.

**Type of School:**
Public: 681 (88%)
Private: 42 (5%)
Charter: 18 (2%)
Home School: 14 (2%)
Parochial: 13 (2%)
Other: 4 (<1%)

**Access to a Computer at Home:**
Yes: 747 (97%)
No: 27 (3%)

**Frequency of Computer Use:**
Almost every day or every day: 438 (57%)
Three or four times per week: 175 (23%)
Once or twice per week: 146 (19%)
Never: 15 (2%)

**Frequency of Internet Use:**
Almost every day or every day: 401 (52%)
Three or four times per week: 189 (24%)
Once or twice per week: 166 (21%)
Never: 18 (2%)

**Computer Classes:**
Yes: 385 (50%)
No: 321 (41%)
Unsure: 68 (9%)

**IEP:**
Yes: 87 (11%) (e.g., ADHD, Dyslexia, Emotional Disturbance, Gifted, Hearing Loss, High Functioning Asperger's, Impaired/Slow Learning, Auditory Processing Disability, Orthopedic Impairment, Speech and Language, Speech Impairment)
No: 631 (82%)
Unsure: 56 (7%)

**Testing Accommodations:**
Yes: 83 (11%) (e.g., Paper Test, Printable Test, Student can take test in another language, ELD, Limited English Proficiency, Listen to questions on tape and use bilingual dictionary, Supervised breaks and additional time, Assessments can be read, Assessments one on one with administrator, Cantonese Bilingual Pathway Instruction, Extra time and modified questions, Extended response time, Separate room)
No: 647 (84%)
Unsure: 42 (5%)

There is no assessment program at this grade level: 1 (<1%)
Child does not participate in the school's testing or assessment program: 1 (<1%)

**ELA Grades:**
Above Average: 375 (48%)
Average: 324 (42%)
Below Average: 51 (7%)
Unsure: 20 (3%)
*Not all participants responded to this question.

**Mathematics Grades:**
Above Average: 392 (51%)
Average: 311 (40%)
Below Average: 55 (7%)
Unsure: 14 (2%)
*Not all participants responded to this question.

**Ethnic/Cultural Breakdown:**
White: 493 (64%)
Hispanic: 137 (18%)
Asian: 125 (16%)
Black/African American: 76 (10%)
Native Hawaiian or Other Pacific Islander: 28 (4%)
American Indian or Alaskan Native: 17 (2%)
Filipino: 12 (2%)
Asian Indian: 5 (<1%)
Other: 3 (<1%)
*Total percentage is over 100% because more than one response could be selected.

**Household Income:**
Under $25,000: 135 (17%)
Between $25,001 and $50,000: 170 (22%)
Between $50,001 and $75,000: 139 (18%)
Between $75,001 and $100,000: 145 (19%)
Between $100,001 and $150,000: 110 (14%)
Over $150,001: 54 (7%)

Appendix G – 	Usability, Accessibility, and Accommodations Guidelines

# Smarter Balanced Assessment Consortium:
## Usability, Accessibility, and Accommodations Guidelines

Prepared with the assistance of the
National Center on Educational Outcomes

August 1, 2014

# Table of Contents

# Introduction

The Smarter Balanced Assessment Consortium (Smarter Balanced) strives to provide every student with a positive and productive assessment experience, generating results that are a fair and accurate estimate of each student's achievement. Further, Smarter Balanced is building on a framework of accessibility for **all** students, including English Language Learners (ELLs), students with disabilities, and ELLs with disabilities, but not limited to those groups. In the process of developing its next-generation assessments to measure students' knowledge and skills as they progress toward college and career readiness, Smarter Balanced recognized that the validity of assessment results depends on each and every student having appropriate universal tools, designated supports, and accommodations when needed based on the constructs being measured by the assessment. This document was developed for the Smarter Balanced member states to guide the selection and administration of universal tools, designated supports, and accommodations.

The Smarter Balanced assessment is based on the Common Core State Standards (CCSS). Thus, the universal tools, designated supports, and accommodations that are appropriate for the Smarter Balanced assessment may be different from those that states allowed in the past. For the secure summative assessments, a state can only make available to students the universal tools, designated supports, and accommodations that are included in the *Smarter Balanced Usability, Accessibility, and Accommodations Guidelines.* A member state may elect **not to make available** to its students, any universal tool, designated support, or accommodation that is otherwise included in the *Guidelines* when the implementation or use of the universal tool, designated support, or accommodation is in conflict with a member state's law, regulation, or policy.

These *Guidelines* describe the Smarter Balanced universal tools, designated supports, and accommodations available for the Smarter Balanced assessments at this time (see Appendix A). The specific universal tools, designated supports, and accommodations approved by Smarter Balanced may change in the future if additional tools, supports or accommodations are identified for the assessment based on state experience and research findings. The Consortium will establish a standing committee, including members from Governing States that will review suggested additional universal tools, designated supports, and accommodations to determine if changes are warranted. Proposed changes to the list of universal tools, designated supports, and accommodations will be brought to Governing States for review, input, and vote for approval. Furthermore, states may issue temporary approvals (i.e., one summative assessment administration) for individual unique student accommodations. State leads will evaluate formal requests for unique accommodations and determine whether or not the request poses a threat to the measurement of the construct. Upon issuing a temporary approval, the State will send documentation of the approval to the Consortium. The Consortium will consider all state approved temporary accommodations as part of the annual Consortium accommodations review process. The Consortium will provide to member states a list of the temporary accommodations issued by states that are not Consortium approved accommodations.

## Intended Audience and Recommended Use

The Smarter Balanced Assessment Consortium's *Usability, Accessibility, and Accommodations Guidelines* are intended for school-level personnel and decision-making teams, particularly Individualized Education Program (IEP) teams, as they prepare for and implement the Smarter Balanced assessment. The Guidelines provide information for classroom teachers, English

development educators, special education teachers, and related services personnel to use in selecting and administering universal tools, designated supports, and accommodations for those students who need them. The Guidelines are also intended for assessment staff and administrators who oversee the decisions that are made in instruction and assessment.

The Smarter Balanced *Guidelines* apply to **all** students. They emphasize an individualized approach to the implementation of assessment practices for those students who have diverse needs and participate in large-scale content assessments. This document focuses on universal tools, designated supports, and accommodations for the Smarter Balanced content assessments of English language arts/literacy and mathematics (math). At the same time, it supports important instructional decisions about accessibility and accommodations for students who participate in the Smarter Balanced assessments. It recognizes the critical connection between accessibility and accommodations in instruction and accessibility and accommodations during assessment. Professional development materials that support the *Guidelines* and this critical instruction-assessment link will be available in the Spring of 2014. The *Guidelines* also are supported by the Smarter Balanced Test Administration Manual.

## Smarter Balanced Assessment Design

The Smarter Balanced Assessment Consortium has developed a system of valid, reliable, and fair next-generation assessments aligned to the CCSS in English language arts (ELA)/literacy and mathematics for grades 3-8 and 11. The system includes summative assessments for accountability purposes, optional interim assessments for local use, and formative tools and processes for instructional use. Computer adaptive testing technologies are used for the summative and interim assessments to provide meaningful feedback and actionable data that teachers and other stakeholders can use to help students succeed. For more information, visit www.smarterbalanced.org/smarter-balanced-assessments/.

## Recognizing Access Needs in All Students

All students (including students with disabilities, ELLs, and ELLs with disabilities) are to be held to the same expectations for participation and performance on state assessments. Specifically, all students enrolled in grades 3-8 and 11 are required to participate in the Smarter Balanced mathematics assessment except:

- Students with the most significant cognitive disabilities who meet the criteria for the mathematics alternate assessment based on alternate achievement standards (approximately 1% or fewer of the student population).

All students enrolled in grades 3-8 and 11 are required to participate in the Smarter Balanced English language/literacy assessment except:

- Students with the most significant cognitive disabilities who meet the criteria for the English language/literacy alternate assessment based on alternate achievement standards (approximately 1% or fewer of the student population).
- ELLs who are enrolled for the first year in a U.S. school. These students instead participate in their state's English language proficiency assessment.

Federal laws governing student participation in statewide assessments include the Elementary and Secondary Education Act (ESEA) (reauthorized as the No Child Left Behind Act of 2001 – NCLB), the

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), and Section 504 of the Rehabilitation Act of 1973 (reauthorized in 2008).

Recognizing the diverse characteristics and needs of students who participate in the Smarter Balanced assessments, the Smarter Balanced states worked together through the Smarter Balanced Test Administration and Student Access Work Group to develop an *Accessibility and Accommodations Framework* that guided the consortium as it worked to reach agreement on the specific tools, supports, and accommodations available for the assessment. The Work Group also considered research-based lessons learned about universal design, accessibility tools, and accommodations (see Appendix B).

The conceptual model that serves as the basis for the *Usability, Accessibility, and Accommodations Guidelines* is shown in Figure 1. This figure portrays several aspects of the Smarter Balanced assessment features – universal tools (available for all students), designated supports (available when indicated by an adult or team), and accommodations (available need is documented in an Individualized Education Program – IEP or 504 plan). It also portrays the additive and sequentially-inclusive nature of these three aspects. Universal tools are available to all students, including those receiving designated supports and those receiving accommodations. Designated supports are available only to students for whom an adult or team has indicated the need for these accommodations (as well as those students for whom the need is documented). Accommodations are available only to those students with documentation of the need through a formal plan (i.e., IEP). Those students also may use designated supports and universal tools.

A universal tool for one content focus may be an accommodation for another content focus (see, for example, calculator). Similarly, a designated support may also be an accommodation, depending on the content target (see, for example, scribe). This approach is consistent with the emphasis that Smarter Balanced has placed on the validity of assessment results coupled with access. **Universal tools, designated supports, and accommodations all yield valid scores that count as participation in statewide assessments when used in a manner consistent with the *Guidelines.***

Also, as shown in Figure 1, for each category of assessment features – universal tools, designated supports, and accommodations – there exist both embedded and non-embedded versions of the tools, supports, or accommodations depending on whether they are provided as digitally-delivered components of the test administration system or separate from it.

Figure 1: Conceptual Model Underlying the Smarter Balanced *Usability, Accessibility, and Accommodations Guidelines*
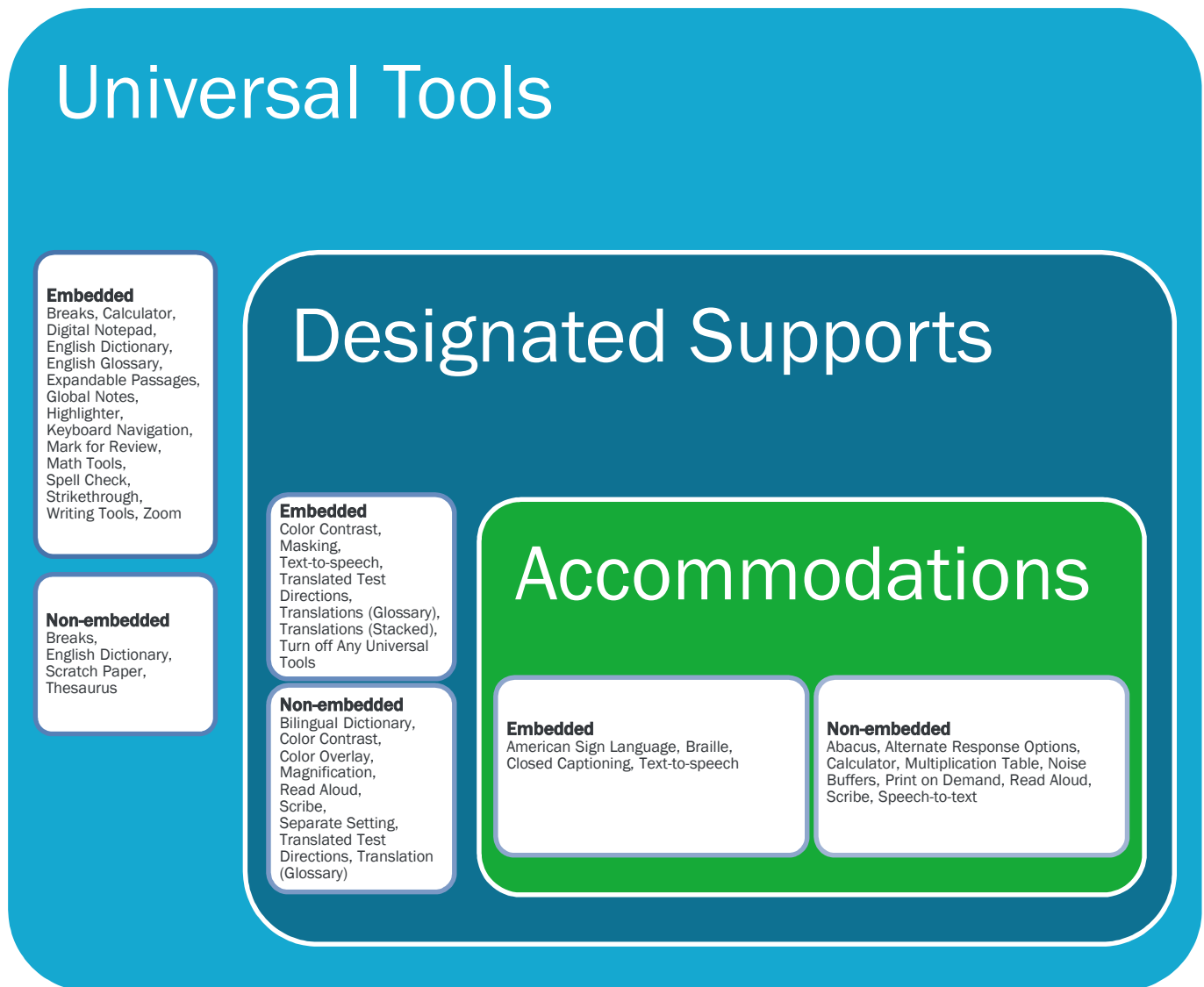


The Conceptual Model recognizes that all students should be held to the same expectations for instruction in CCSS and have available to them universal accessibility features. It also recognizes that some students may have certain characteristics and access needs that require the use of accommodations for instruction and when they participate in the Smarter Balanced assessments.

These *Guidelines* present the current universal tools, designated supports, and accommodations adopted by the Smarter Balanced states to ensure valid assessment results for all students taking its assessments.

## Structure of This Document

This document is divided into several parts:

- **Introduction:** This section introduces the document and the conceptual model that is the basis for the universal tools, designated supports, and accommodations in the *Guidelines*.

- **Section I:** This section features the Consortium's universal tools.

- **Section II:** This section features the designated supports available on Smarter Balanced assessments.

- **Section III:** This section features the accommodations available on Smarter Balanced assessments.

- **Appendix A:** This appendix provides a summary list of Smarter Balanced's universal tools, designated supports, and accommodations.

- **Appendix B:** This appendix describes lessons learned from research on universal design, accessibility tools, and accommodations.

# Usability, Accessibility, and Accommodations Guidelines

# Section I: Smarter Balanced Universal Tools

## What Are Universal Tools?

**Universal tools** are access features of the assessment that are either provided as digitally-delivered components of the test administration system or separate from it. Universal tools are available to all students based on student preference and selection.

## Embedded Universal Tools

The Smarter Balanced digitally-delivered assessments include a wide array of embedded universal tools. These are available to all students as part of the technology platform.

Table 1 lists the embedded universal tools available to all students for computer administered Smarter Balanced assessments. It includes a description of each tool. Although these tools are generally available to all students, educators may determine that one or more might be distracting for a particular student, and thus might indicate that the tool should be turned off for the administration of the assessment to the student (see Section II – Designated Supports).

Table 1. Embedded Universal Tools Available to All Students

| Universal Tool | Description |
|---|---|
| Breaks | The number of items per session can be flexibly defined based on the student's need. Breaks of more than 20 minutes will prevent the student from returning to items already attempted by the student. There is no limit on the number of breaks that a student might be given. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |
| Calculator (for calculator-allowed items only) (See Non-embedded Accommodations for students who cannot use the embedded calculator) | An embedded on-screen digital calculator can be accessed for calculator-allowed items when students click on the calculator button. This tool is available only with the specific items for which the Smarter Balanced Item Specifications indicated that it would be appropriate. When the embedded calculator, as presented for all students, is not appropriate for a student (for example, for a student who is blind), the student may use the calculator offered with assistive technology devices (such as a talking calculator or a braille calculator). |
| Digital notepad | This tool is used for making notes about an item. The digital notepad is item-specific and is available through the end of the test segment. Notes are not saved when the student moves on to the next segment or after a break of more than 20 minutes. |
| English Dictionary (for ELA-performance task full writes) | An English dictionary **may** be available for the full write portion of an ELA performance task, **pending contractual discussions**. A full write is the second part of a performance task. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |
| English glossary | Grade- and context-appropriate definitions of specific construct-irrelevant terms are shown in English on the screen via a pop-up window. The student can access the embedded glossary by clicking on any of the pre-selected terms. The use of this accommodation may result in the student needing additional overall time to complete the assessment. |

| Universal Tool | Description |
|---|---|
| Expandable passages | Each passage or stimulus can be expanded so that it takes up a larger portion of the screen. |
| Global notes (for ELA performance tasks) | Global notes is a notepad that is available for ELA performance tasks in which students complete a full write. A full write is the second part of a performance task. The student clicks on the notepad icon for the notepad to appear. During the ELA performance tasks, the notes are retained from segment to segment so that the student may go back to the notes even though the student is not able to go back to specific items in the previous segment. |
| Highlighter | A digital tool for marking desired text, item questions, item answers, or parts of these with a color. Highlighted text remains available throughout each test segment. |
| Keyboard navigation | Navigation throughout text can be accomplished by using a keyboard. |
| Mark for review | Allows students to flag items for future review during the assessment. Markings are not saved when the student moves on to the next segment or after a break of more than 20 minutes. |
| Math tools | These digital tools (i.e., embedded ruler, embedded protractor) are used for measurements related to math items. They are available only with the specific items for which the *Smarter Balanced Item Specifications* indicate that one or more of these tools would be appropriate. |
| Spell check (for ELA items) | Writing tool for checking the spelling of words in student-generated responses. Spell check only gives an indication that a word is misspelled; it does not provide the correct spelling. This tool is available only with the specific items for which the Smarter Balanced Item Specifications indicated that it would be appropriate. Spell check is bundled with other embedded writing tools for all performance task full writes (planning, drafting, revising, and editing). A full write is the second part of a performance task. |
| Strikethrough | Allows users to cross out answer options. If an answer option is an image, a strikethrough line will not appear, but the image will be grayed out. |
| Writing tools | Selected writing tools (i.e., bold, italic, bullets, undo/redo) are available for all student-generated responses. (Also see spell check.) |
| Zoom | A tool for making text or other graphics in a window or frame appear larger on the screen. The default font size for all tests is 14 pt. The student can make text and graphics larger by clicking the *Zoom In* button. The student can click the *Zoom Out* button to return to the default or smaller print size. When using the zoom feature, the student only changes the size of text and graphics on the current screen. To increase the default print size of the entire test (from 1.5X to 3.0X default size), the print size must be set for the student in the Test Information Distribution Engine (TIDE, or state's comparable platform), or set by the test administrator prior to the start of the test. This is the only feature that test administrators can set. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |

## Non-embedded Universal Tools

Some universal tools may need to be provided outside of the computer test administration system. These tools, shown in Table 2, are to be provided locally for those students. They can be made available to any student.

Table 2. Non-embedded Universal Tools Available to All Students

| Universal Tool | Description |
|---|---|
| Breaks | Breaks may be given at predetermined intervals or after completion of sections of the assessment for students taking a paper-based test. Sometimes students are allowed to take breaks when individually needed to reduce cognitive fatigue when they experience heavy assessment demands. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |
| English Dictionary (for ELA-performance task full writes) | An English dictionary can be provided for the full write portion of an ELA performance task. A full write is the second part of a performance task. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |
| Scratch paper | Scratch paper to make notes, write computations, or record responses may be made available. Only plain paper or lined paper is appropriate for ELA. Graph paper is required beginning in sixth grade and can be used on all math assessments. A student can use an assistive technology device for scratch paper as long as the device is certified.[1]<br><br>**CAT:** All scratch paper must be collected and securely destroyed at the end of each CAT assessment session to maintain test security.<br><br>**Performance Tasks:** For mathematics and ELA performance tasks, if a student needs to take the performance task in more than one session, scratch paper may be collected at the end of each session, securely stored, and made available to the student at the next performance task testing session. Once the student completes the performance task, the scratch paper must be collected and securely destroyed to maintain test security. |
| Thesaurus (for ELA-performance task full writes) | A thesaurus contains synonyms of terms while a student interacts with text included in the assessment. A full write is the second part of a performance task. The use of this universal tool may result in the student needing additional overall time to complete the assessment. |

Appendix A provides a summary of universal tools, designated supports, and accommodations (both embedded and non-embedded) available for the Smarter Balanced assessments.

---

[1] Smarter Balanced is working closely with our test administration platform vendor to create a process through which assistive technology devices can be certified. Certification ensures that the device functions properly and appropriately addresses test security.

# Section II: Smarter Balanced Designated Supports

## What Are Designated Supports?

**Designated supports** for the Smarter Balanced assessments are those features that are available for use by **any student** for whom the need has been indicated by an educator (or team of educators with parent/guardian and student). Scores achieved by students using designated supports will be included for federal accountability purposes. It is recommended that a consistent process be used to determine these supports for individual students. All educators making these decisions should be trained on the process and should be made aware of the range of designated supports available. Smarter Balanced states have identified digitally-embedded and non-embedded designated supports for students for whom an adult or team has indicated a need for the support.

Designated supports need to be identified prior to assessment administration. Embedded and non-embedded supports must be entered into the Test Information Distribution Engine (TIDE, or state's comparable platform). Any non-embedded designated supports must be acquired prior to testing.

## Who Makes Decisions About Designated Supports?

Informed adults make decisions about designated supports. Ideally, the decisions are made by all educators familiar with the student's characteristics and needs, as well as those supports that the student has been using during instruction and for other assessments. Student input to the decision, particularly for older students, is also recommended.

Forthcoming professional development materials to be available through Smarter Balanced will provide suggestions of processes that may be used if a district or school does not have an existing process in place for adults and others to make decisions about designated supports. The use of an *Individual Student Assessment Accessibility Profile (ISAAP)*, created and provided by Smarter Balanced, is one process that may be used to determine which designated supports should be available for an individual student. Schools may choose to use another decision-making process. **Regardless of the process used, all embedded designated supports must be activated prior to testing by entering information in the TIDE, or state's comparable platform.**

## Embedded Designated Supports

Table 3 lists the embedded designated supports available to all students for whom the need has been indicated. It includes a description of each support along with recommendations for when the support might be needed.

Table 3. Embedded Designated Supports

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| Color contrast | Enable students to adjust screen background or font color, based on student needs or preferences. This may include reversing the colors for the entire interface or choosing the color of font and background. | Students with attention difficulties may need this support for viewing test content. It also may be needed by some students with visual impairments or other print disabilities (including learning disabilities). Choice of colors should be |

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| | | informed by evidence that color selections meet the student's needs. |
| Masking | Masking involves blocking off content that is not of immediate need or that may be distracting to the student. Students are able to focus their attention on a specific part of a test item by masking. | Students with attention difficulties may need to mask content not of immediate need or that may be distracting during the assessment. This support also may be needed by students with print disabilities (including learning disabilities) or visual impairments. Masking allows students to hide and reveal individual answer options, as well as all navigational buttons and menus. |
| Text-to-speech (for math stimuli items and ELA items, not for reading passages)[2] (See Embedded Accommodations for ELA reading passages) | Text is read aloud to the student via embedded text-to-speech technology. The student is able to control the speed as well as raise or lower the volume of the voice via a volume control. | Students who are struggling readers may need assistance accessing the assessment by having all or portions of the assessment read aloud. This support also may be needed by students with reading-related disabilities, or by students who are blind and do not yet have adequate braille skills. This support will likely be confusing and may impede the performance of students who do not regularly have the support during instruction. Students who use text-to-speech will need headphones unless tested individually in a separate setting. |
| Translated test directions (for math items) | Translation of test directions is a language support available prior to beginning the actual test items. Students can see test directions in another language. | Students who have limited English language skills can use the translated directions support. This support should only be used for students who are proficient readers in the other language and not proficient in English. |
| Translations (glossaries) (for math items) | Translated glossaries are a language support. The translated glossaries are provided for selected construct-irrelevant terms for math. Translations for these terms appear on the computer screen when students click on them. Students with the language glossary setting enabled can view the translated glossary. Students can also select the audio icon next to the glossary term and listen to the audio recording of the glossary. | Students who have limited English language skills (whether or not designated as ELLs or ELLs with disabilities) can use the translation glossary for specific items. The use of this support may result in the student needing additional overall time to complete the assessment. |

---

[2] See Embedded Accommodations for guidelines on the use of Text-to-speech for ELA reading passages.

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| Translations (stacked) (for math items) | Stacked translations are a language support. Stacked translations are available for some students; stacked translations provide the full translation of each test item above the original item in English. | For students whose primary language is not English and who use dual language supports in the classroom, use of the stacked (dual language) translation may be appropriate. Students participate in the assessment regardless of the language. This support will increase reading load and cognitive load. The use of this support may result in the student needing additional overall time to complete the assessment. |
| Turn off any universal tools | Disabling any universal tools that might be distracting or that students do not need to use, or are unable to use. | Students who are easily distracted (whether or not designated as having attention difficulties or disabilities) may be overwhelmed by some of the universal tools. Knowing which specific tools may be distracting is important for determining which tools to turn off. |

## Non-embedded Designated Supports

Some designated supports may need to be provided outside of the digital-delivery system. These supports, shown in Table 4, are to be provided locally for those students unable to use the designated supports when provided digitally.

Table 4. Non-embedded Designated Supports

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| Bilingual dictionary (for ELA-performance task full writes) | A bilingual/dual language word-to-word dictionary is a language support. A bilingual/dual language word-to-word dictionary can be provided for the full write portion of an ELA performance task. A full write is the second part of a performance task. | For students whose primary language is not English and who use dual language supports in the classroom, use of a bilingual/dual language word-to-word dictionary may be appropriate. Students participate in the assessment regardless of the language. The use of this support may result in the student needing additional overall time to complete the assessment. |
| Color contrast | Test content of online items may be printed with different colors. | Students with attention difficulties may need this support for viewing the test when digitally-provided color contrasts do not meet their needs. Some students with visual impairments or other print disabilities (including learning disabilities) also may need this support. Choice of colors should be informed by evidence of |

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| | | those colors that meet the student's needs. |
| Color overlays | Color transparencies are placed over a paper-based assessment. | Students with attention difficulties may need this support to view test content. This support also may be needed by some students with visual impairments or other print disabilities (including learning disabilities). Choice of color should be informed by evidence of those colors that meet the student's needs. |
| Magnification | The size of specific areas of the screen (e.g., text, formulas, tables, graphics, and navigation buttons) may be adjusted by the student with an assistive technology device. Magnification allows increasing the size to a level not provided for by the Zoom universal tool. | Students used to viewing enlarged text or graphics, or navigation buttons may need magnification to comfortably view content. This support also may meet the needs of students with visual impairments and other print disabilities. The use of this designated support may result in the student needing additional overall time to complete the assessment. |
| Read aloud (for math items and ELA items, not for reading passages) (See Non-embedded Accommodations for ELA reading passages) | Text is read aloud to the student by a trained and qualified human reader who follows the administration guidelines provided in the *Smarter Balanced Test Administration Manual*. All or portions of the content may be read aloud. | Students who are struggling readers may need assistance accessing the assessment by having all or portions of the assessment read aloud. This support also may be needed by students with reading-related disabilities, or by students who are blind and do not yet have adequate braille skills. If not used regularly during instruction, this support is likely to be confusing and may impede the performance on assessments. Readers should be provided to students on an individual basis – not to a group of students. A student should have the option of asking a reader to slow down or repeat text. The use of this support may result in the student needing additional overall time to complete the assessment. |
| Scribe (for ELA non-writing items and math items)[3] | Students dictate their responses to a human who records verbatim what they dictate. The scribe must be trained and qualified, and must follow the administration guidelines provided in the | Students who have documented significant motor or processing difficulties, or who have had a recent injury (such as a broken hand or arm) that make it difficult to produce responses may need to dictate their |

---

[3] See Accommodations for use of Scribe for Writing items

| Designated Support | Description | Recommendations for Use |
|---|---|---|
| (See Accommodations for Writing) | *Smarter Balanced Test Administration Manual.* | responses to a human, who then records the students' responses verbatim. The use of this support may result in the student needing additional overall time to complete the assessment. |
| Separate setting | Test location is altered so that the student is tested in a setting different from that made available for most students. | Students who are easily distracted (or may distract others) in the presence of other students, for example, may need an alternate location to be able to take the assessment. The separate setting may be in a different room that allows them to work individually or among a smaller group, or in the same room but in a specific location (for example, away from windows, doors, or pencil sharpeners, in a study carrel, near the teacher's desk, or in the front of a classroom). Some students may benefit from being in an environment that allows for movement, such as being able to walk around. In some instances, students may need to interact with instructional or test content outside of school, such as in a hospital or their home. A specific adult, trained in a manner consistent with the TAM, can act as test proctor (test administrator) when student requires it. |
| Translated test directions | PDF of directions translated in each of the languages currently supported. Bilingual adult can read to student. | Students who have limited English language skills (whether or not designated as ELLs or ELLs with disabilities) can use the translated test directions. In addition, a biliterate adult trained in the test administration manual can read the test directions to the student. The use of this support may result in the student needing additional overall time to complete the assessment. |
| Translations (glossaries) (for math items) | Translated glossaries are a language support. Translated glossaries are provided for selected construct-irrelevant terms for math. Glossary terms are listed by item and include the English term and its translated equivalent. | Students who have limited English language skills can use the translation glossary for specific items. The use of this support may result in the student needing additional overall time to complete the assessment. |

Appendix A provides a summary of universal tools, designated supports, and accommodations (both embedded and non-embedded) available for the Smarter Balanced assessments.

# Section III: Smarter Balanced Accommodations

## What Are Accommodations?

**Accommodations** are changes in procedures or materials that increase equitable access during the Smarter Balanced assessments. Assessment accommodations generate valid assessment results for students who need them; they allow these students to show what they know and can do. Smarter Balanced states have identified digitally-embedded and non-embedded **accommodations** for students for whom there is documentation of the need for the accommodations on an Individualized Education Program (IEP) or 504 accommodation plan. One exception to the IEP or 504 requirement is for students who have had a physical injury (e.g., broken hand or arm) that impairs their ability to use a computer. These students may use the speech-to-text or the scribe accommodations (if they have had sufficient experience with the use of these), as noted in this section.

Determination of which accommodations an individual student will have available for the assessment is necessary because these accommodations must be made available before the assessment, either by entering information into the TIDE, or state's comparable platform, for embedded accommodations, or by ensuring that the materials or setting are available for the assessment for non-embedded accommodations.

The Smarter Balanced Test Administration and Student Access Workgroup recognized that accommodations could increase cognitive load or create other challenges for students who do not need them or who have not had experience using them. Because of this possibility, Smarter Balanced states agreed that **a student's parent/guardian should know about the availability of specific accommodations through a parent/guardian report. This would ensure that parents/guardians are aware of the conditions under which their child participated in the assessment. Information included in the parent/guardian report should not be the basis for any educational decisions (such as eligibility for an Advanced Placement class) nor for documenting/reporting the use of the accommodation elsewhere (such as on a transcript).**

## Who Makes Decisions About Accommodations?

IEP teams and educators make decisions about accommodations. These teams (or educators for 504 plans) provide evidence of the need for accommodations and ensure that they are noted on the IEP or 504 plan.

**The IEP team (or educator developing the 504 plan) is responsible for ensuring that information from the IEP is entered into the TIDE, or state's comparable platform, so that all embedded accommodations can be activated prior to testing.** This can be accomplished by identifying one person from the team to enter information into the TIDE, or state's comparable platform, or by providing information to the test coordinator who enters into the TIDE, or state's comparable platform, a form that lists all accommodations and designated supports needed by individual students on IEPs or 504 plans.

## Embedded Accommodations

Table 5 lists the embedded accommodations available for the Smarter Balanced assessments for those students for whom the accommodations are included on an IEP or 504 plan. The table includes a description of each accommodation along with recommendations for when the accommodation might be needed and how it can be used. For those accommodations that may be considered controversial, a description of considerations about the use of the accommodation is provided.

Table 5. Embedded Accommodations

| Accommodation | Description | Recommendations for Use |
|---|---|---|
| American Sign Language (ASL)<br><br>(for ELA Listening items and math items) | Test content is translated into ASL video. ASL human signer and the signed test content are viewed on the same screen. Students may view portions of the ASL video as often as needed. | Some students who are deaf or hard of hearing and who typically use ASL may need this accommodation when accessing text-based content in the assessment. The use of this accommodation may result in the student needing additional overall time to complete the assessment. For many students who are deaf or hard of hearing, viewing signs is the only way to access information presented orally. It is important to note, however, that some students who are hard of hearing will be able to listen to information presented orally if provided with appropriate amplification and a setting in which extraneous sounds do not interfere with clear presentation of the audio presentation in a listening test. |
| Braille | A raised-dot code that individuals read with the fingertips. Graphic material (e.g., maps, charts, graphs, diagrams, and illustrations) is presented in a raised format (paper or thermoform). Contracted and non-contracted braille is available; Nemeth code is available for math. | Students with visual impairments may read text via braille. Tactile overlays and graphics also may be used to assist the student in accessing content through touch. Refreshable braille is available only for ELA because Nemeth Code is not available via refreshable braille. For math, braille will be presented via embosser; embosser-created braille can be used for ELA also. The type of braille presented to the student (contracted or non-contracted) is set in TIDE, or state's comparable platform. The use of this accommodation may result in the student needing additional overall time to complete the assessment. |
| Closed captioning<br><br>(for ELA listening items) | Printed text that appears on the computer screen as audio materials are presented. | Students who are deaf or hard of hearing and who typically access information presented via audio by reading words that appear in synchrony with the audio |

| Accommodation | Description | Recommendations for Use |
|---|---|---|
| | | presentation may need this support to access audio content. For many students who are deaf or hard of hearing, viewing words (sometimes in combination with reading lips and ASL) is how they access information presented orally. It is important to note, however, that some students who are hard of hearing will be able to listen to information presented orally if provided with appropriate amplification and a setting in which extraneous sounds do not interfere with clear presentation of the audio presentation in a listening test. |
| Text-to-speech (for ELA reading passages) | Text is read aloud to the student via embedded text-to-speech technology. The student is able to control the speed as well as raise or lower the volume of the voice via a volume control. | **This accommodation is appropriate for a very small number of students (estimated to be approximately 1-2% of students with disabilities participating in a general assessment).**<br><br>• For students in grades 3 - 5, text-to-speech will not be an available accommodation. Content experts agree that this accommodation should not be provided during these grades because it would compromise the construct being measured.<br><br>• For students in grades 6 – 8 and 11, text-to-speech is available as an accommodation for students whose need is documented in an IEP or 504 plan.<br><br>Reports can be run to indicate the percent of students who had access to text-to-speech on reading test passages.<br><br>Students who use text-to-speech will need headphones unless tested individually in a separate setting. |

## Non-embedded Accommodations

Table 6 lists the non-embedded accommodations available for the Smarter Balanced assessments for those students for whom the accommodations are documented on an IEP or 504 plan. The table includes a description of each accommodation, along with recommendations for when the accommodation might be needed and how it can be used. For those accommodations that may be considered controversial, a description of considerations about the use of the accommodation is provided.

Table 6. Non-embedded Accommodations Available

| Accommodation | Description | Recommendations for Use |
|---|---|---|
| Abacus | This tool may be used in place of scratch paper for students who typically use an abacus. | Some students with visual impairments who typically use an abacus may use an abacus in place of using scratch paper. |
| Alternate response options | Alternate response options include but are not limited to adapted keyboards, large keyboards, StickyKeys, MouseKeys, FilterKeys, adapted mouse, touch screen, head wand, and switches. | Students with some physical disabilities (including both fine motor and gross motor skills) may need to use the alternate response options accommodation. Some alternate response options are external devices that must be plugged in and be compatible with the assessment delivery platform. |
| Calculator (for calculator allowed items only) | A non-embedded calculator for students needing a special calculator, such as a braille calculator or a talking calculator, currently unavailable within the assessment platform. | Students with visual impairments who are unable to use the embedded calculator for calculator-allowed items will be able to use the calculator that they typically use, such as a braille calculator or a talking calculator. Test administrators should ensure that the calculator is available only for designated calculator items. |
| Multiplication Table (grade 4 and above math items) | A paper-based single digit (1-9) multiplication table will be available from Smarter Balanced for reference. | For students with a documented and persistent calculation disability (i.e., dyscalculia). |
| Noise Buffers | Ear mufflers, white noise, and/or other equipment used to block external sounds. | Student (not groups of students) wears equipment to reduce environmental noises. Students may have these testing variations if regularly used in the classroom. Students who use noise buffers will need headphones unless tested individually in a separate setting. |
| Print on demand | Paper copies of either passages/stimuli and/or items are | Some students with disabilities may need paper copies of either passages/stimuli |

| Accommodation | Description | Recommendations for Use |
|---|---|---|
| | printed for students. For those students needing a paper copy of a passage or stimulus, permission for the students to request printing must first be set in TIDE, or state's comparable platform. For those students needing a paper copy of one or more items, the Smarter Balanced Help Desk (1-855-833-1969) must be contacted by the school or district coordinator to have the accommodation set for the student. | and/or items. A very small percentage of students should need this accommodation. The use of this accommodation may result in the student needing additional time to complete the assessment. |
| Read aloud<br><br>(for ELA reading **passages**, grades 6-8 and 11; **blind students in grades 3-8 and 11 who do not yet have adequate braille skills**) | Text is read aloud to the student by a trained and qualified human reader who follows the administration guidelines provided in the *Smarter Balanced Test Administration Manual*. All or portions of the content may be read aloud. | **This accommodation is appropriate for a very small number of students (estimated to be approximately 1-2% of students with disabilities participating in a general assessment).**<br><br>• For students in grades 3 - 5, read aloud will not be an available accommodation. Content experts agree that this accommodation should not be provided during these grades because it would compromise the construct being measured.<br><br>• For students in grades 6 – 8 and 11, read aloud is available as an accommodation for students whose need is documented in an IEP or 504 plan.<br><br>Reports can be run to indicate the percent of students who had access to read aloud on reading test passages.<br><br>Readers should be provided to students on an individual basis – not to a group of students. A student should have the option of asking a reader to slow down or repeat text. |
| Scribe<br><br>(See Designated Supports for math and non-writing ELA) | Students dictate their responses to a human who records verbatim what they dictate. The scribe must be trained and qualified, and must follow the administration guidelines provided in the *Smarter Balanced Test Administration Manual.* | Students who have documented significant motor or processing difficulties, or who have had a recent injury (such as a broken hand or arm) that makes it difficult to produce responses may need to dictate their responses to a human, who then records the students' responses verbatim. The use of this accommodation may result in the student |

| Accommodation | Description | Recommendations for Use |
|---|---|---|
| | | needing overall additional time to complete the assessment. For many of these students, dictating to a human scribe is the only way to demonstrate their composition skills. It is important that these students be able to develop planning notes via the human scribe, and to view what they produce while composing via dictation to the scribe. |
| Speech-to-text | Voice recognition allows students to use their voices as input devices to the computer, to dictate responses or give commands (e.g., opening application programs, pulling down menus, and saving work). Voice recognition software generally can recognize speech up to 160 words per minute. Students may use their own assistive technology devices. | Students who have motor or processing disabilities (such as dyslexia) or who have had a recent injury (such as a broken hand or arm) that make it difficult to produce text or commands using computer keys may need alternative ways to work with computers. Students will need to be familiar with the software, and have had many opportunities to use it prior to testing. Speech-to-text software requires that the student go back through all generated text to correct errors in transcription, including use of writing conventions; thus, prior experience with this accommodation is essential. If students use their own assistive technology devices, all assessment content should be deleted from these devices after the test for security purposes. For many of these students, using voice recognition software is the only way to demonstrate their composition skills. Still, use of speech-to-text does require that students know writing conventions and that they have the review and editing skills required of students who enter text via the computer keyboard. It is important that students who use speech-to-text also be able to develop planning notes via speech-to-text, and to view what they produce while composing via speech-to-text. |

# Resources

Christensen, L., Carver, W., VanDeZande, J., & Lazarus, S. (2011). *Accommodations manual: How to select, administer, and evaluate the use of accommodations for instruction and assessment of students with disabilities (3rd ed.).* Washington, DC: Assessing Special Education Students State Collaborative on Assessment and Student Standards, Council of Chief State School Officers.

Christensen, L., Shyyan, V., Schuster, T., Mahaley, P., & Saez, S. (2012). *Accommodations manual: How to select, administer, and evaluate use of accommodations for instruction and assessment of English language learners.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Fedorchak, G. (2012). *Access by Design – Implications for equity and excellence in education*. Draft paper prepared for the Smarter Balanced Assessment Consortium.

Measured Progress. (2013). *Framework for Accessibility and Accommodations.* Smarter Balanced Assessment Consortium. (Forthcoming Spring 2014)

National Center on Educational Outcomes. (2009). *Accommodations bibliography.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at: https://apps.cehd.umn.edu/nceo/accommodations/

National Council on Measurement in Education. (2012). *Testing and data integrity in the administration of statewide student assessment programs.*

*Professional Development Module.* (Forthcoming Spring 2014)

Shyyan, V., Christensen, L., Touchette, B., Lightborne, L., Gholson, M., & Burton, K. (2013). *Accommodations manual: How to select, administer, and evaluate use of accommodations for instruction and assessment of English language learners with disabilities.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Smarter Balanced. (2012). *Translation accommodations framework for testing ELLs in mathematics.* Available at: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/Translation-Accommodations-Framework-for-Testing-ELL-Math.pdf

Smarter Balanced. (2012). *Accommodations for English Language Learners and Students with Disabilities: A research-based decision algorithm*. Available at: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-represented-students.pdf

# Appendix A: Summary of Smarter Balanced Universal Tools, Designated Supports, and Accommodations

| | Universal Tools | Designated Supports | Accommodations |
|---|---|---|---|
| Embedded | Breaks<br>Calculator[1]<br>Digital Notepad<br>English Dictionary[2]<br>English Glossary<br>Expandable Passages<br>Global Notes<br>Highlighter<br>Keyboard Navigation<br>Mark for Review<br>Math Tools[3]<br>Spell Check[4]<br>Strikethrough<br>Writing Tools[5]<br>Zoom | Color Contrast<br>Masking<br>Text-to-Speech[6]<br>Translated Test Directions[7]<br>Translations (Glossary)[8]<br>Translations (Stacked)[9]<br>Turn off Any Universal Tools | American Sign Language[10]<br>Braille<br>Closed Captioning[11]<br>Text-to-Speech[12] |
| Non-embedded | Breaks<br>English Dictionary[13]<br>Scratch Paper<br>Thesaurus[14] | Bilingual Dictionary[15]<br>Color Contrast<br>Color Overlay<br>Magnification<br>Read Aloud<br>Scribe[16]<br>Separate Setting<br>Translated Test Directions<br>Translations (Glossary)[17] | Abacus<br>Alternate Response Options[18]<br>Calculator[19]<br>Multiplication Table[20]<br>Noise Buffers<br>Print on Demand<br>Read Aloud<br>Scribe<br>Speech-to-Text |

*Items shown are available for ELA and Math unless otherwise noted.

[1] For calculator-allowed items only

[2] For ELA performance task full-writes

[3] Includes embedded ruler, embedded protractor

[4] For ELA items

[5] Includes bold, italic, underline, indent, cut, paste, spell check, bullets, undo/redo.

[6] For ELA items (not ELA reading passages) and math items

[7] For math items

[8] For math items

[9] For math test

[10] For ELA listening Items and math items

[11] For ELA listening items

[12] For ELA reading passages grades 6-8 and 11

[13] For ELA performance task full-writes

[14] For ELA performance task full-writes

[15] For ELA performance task full-writes

[16] For ELA non-writing items and math items

[17] For math items

[18] Includes adapted keyboards, large keyboards, StickyKeys, MouseKeys, FilterKeys, adapted mouse, touch screen, head wand, and switches.

[19] For calculator-allowed items only

[20] For math items beginning in grade 4.

# Appendix B: Research-based Lessons Learned about Universal Design, Accessibility Tools, and Accommodations

More than half of all states in the United States participated in research spurred by the opportunity that states had to develop alternate assessments based on modified achievement standards (AA-MAS). The research conducted since 2007 provides numerous findings that are relevant to the next generation assessments. Lessons learned from this research that are relevant to the Smarter Balanced assessment system are highlighted here[21]

## Who might benefit from accessibility features identified by AA-MAS research?

Several studies explored the characteristics of students who might benefit from an AA-MAS and the accessibility features incorporated in the assessment. These studies consistently found:

- Students with and without Individualized Education Programs (IEPs) and 504 plans would likely benefit from assessments with increased accessibility features.
- Students identified for the AA-MAS or who were among the lowest performing students in a state tended to be males, ethnic or racial minorities, English language learners, or from low socioeconomic backgrounds.
- Students identified for the AA-MAS tended to have difficulty with:
    - Print materials
    - High vocabulary load materials
    - Directions
    - Multi-step problem solving
- Students identified for the AA-MAS tended to have:
    - Distractibility
    - Limited meta-cognitive skills
    - Poor organizational skills
    - Poor self-monitoring skills
    - Slower work pace
    - Limited working memory capacity

## What changes can be made to test items and tests that do not change the construct being assessed?

Many studies examined the effects of changes to test items or the tests themselves. Among those changes that did not violate the construct were:

- Enhanced directions
- Increased size of text and visuals
- Increased white space
- Simplified formats, including simplified visuals
- Underlining

---

[21] The research used to develop this summary was highlighted in the document *Lessons Learned in Federally Funded Projects That Can Improve the Instruction and Assessment of Low Performing Students with Disabilities,* edited by M. Thurlow, S. Lazarus, and S. Bechard (2012), available at www.nceo.info/OnlinePubs/LessonsLearned.pdf, and presentations by the authors of three of the chapters in the *Lessons Learned* report, Sue Bechard, Vince Dean, Sheryl Lazarus, and Shelly Loving-Ryder, along with representatives from the two general assessment consortia (PARCC – Tamara Reavis; Smarter Balanced – Magda Chia).

Among those changes that might not violate the construct, depending on how the construct was specifically defined, were:

- Adding visuals
- Bolding text
- Simplifying language in item stems
- Changing distractors by editing the attractive distractor or changing the order of distractors
- Chunking text by embedding questions within a passage
- Reordering items
- Providing thought questions or hint boxes
- Scaffolding for vocabulary, definition, context, inference, or complex questions

Other findings highlighted the need for individualized decisions about some accessibility features. For example:

- Read-aloud features are differentially effective for and preferred by students
- Some features increase engagement and motivation in students
- Too many features can be confusing to students

Researchers found that students needed to have the opportunity to practice new item types and new accessibility features. In addition, their research emphasized the benefits of cognitive labs and item tryouts with students.

## What can test developers do to build on the lessons learned from AA-MAS research and implementation?

Many studies and AA-MAS implementation efforts pointed to considerations for test developers. For example:
- Require item-writer training that focuses on universal design and accessibility principles
- Develop items from scratch rather than attempting to modify existing items to increase universal design and accessibility characteristics
- Ensure that all users understand the purpose of the assessment through professional development activities
- Always consider format changes that might increase the accessibility of items and tests, but make changes to content and cognitive load only after careful delineation of the purpose and content targets of the assessment.
- Engage in research on the effects of individual changes and combinations of changes intended to increase universal design and accessibility.
- Implement innovative items with caution, and only after exploring the accessibility implications of the innovative items.

# Appendix C: Frequently Asked Questions

Smarter Balanced states identified frequently asked questions (FAQs) and developed applicable responses to support the information provided in the Smarter Balanced Assessment Consortium's *Usability, Accessibility, and Accommodations Guidelines*. These questions and responses, as well as the information in the *Guidelines* document apply to the Smarter Balanced interim and summative assessments.

States may use these FAQs to assist districts and schools with transitioning from their former assessments to the Smarter Balanced assessments. In addition, the FAQs may be used by districts to ensure understanding among staff and schools regarding the universal tools, designated supports, and accommodations available for the Smarter Balanced assessments. Schools may use them with decision-making teams (including parents) as decisions are made and implemented with respect to use of the Smarter Balanced *Usability, Accessibility, and Accommodations Guidelines.*

Additional information to aid in the implementation of the *Guidelines* is available in the *Individual Student Assessment Accessibility Profile (ISAAP) Module*, the *Test Administration Manual*, and the *Implementation Guide*. These documents will be made available over the next few weeks.

The FAQs are organized into four sections. First are general questions. Second is a set of questions about specific universal tools and designated supports. Questions that pertain specifically to English language learners (ELLs) comprise the third set of FAQs, and questions that pertain specifically to students with disabilities comprise the fourth set of FAQs.

## General FAQs

1. *What are the differences among the three categories of universal tools, designated supports, and accommodations?*

   **Universal tools** are access features that are available to all students based on student preference and selection. **Designated supports** for the Smarter Balanced assessments are those features that are available for use by any student (including English language learners, students with disabilities, and English language learners with disabilities) for whom the need has been indicated by an educator or team of educators (with parent/guardian and student input as appropriate). **Accommodations** are changes in procedures or materials that increase equitable access during the Smarter Balanced assessments by generating valid assessment results for students who need them and allowing these students the opportunity to show what they know and can do. The *Usability, Accessibility, and Accommodations Guidelines* identify accommodations for students for whom there is documentation of the need for the accommodations on an Individualized Education Program (IEP) or 504 accommodation plan.

   Universal tools, designated supports, and accommodations may be either embedded in the test administration system or provided locally (non-embedded).

2. *Which students should use each category of universal tools, designated supports, and accommodations?*

> **Universal tools** are available to all students, including those receiving designated supports and those receiving accommodations. Designated supports are available only to students for whom an adult or team (consistent with state-designated practices) has indicated the need for these supports (as well as those students for whom the need is documented). **Accommodations** are available only to those students with documentation of the need through either an Individualized Education Program (IEP) or a 504 accommodation plan. Students who have IEPs or 504 accommodation plans also may use designated supports and universal tools.

**What Tools Are Available for my Student?**

|  | All Students | English language learners (ELLs) | Students with disabilities | ELLs with disabilities |
|---|---|---|---|---|
| Universal Tools |  |  |  |  |
| Designated Supports | 1 | 1 |  |  |
| Accommodations |  |  |  |  |

*1 Only for instances that an adult (or team) has deemed the supports appropriate for a specific student's testing needs.*

3. *What is the difference between embedded and non-embedded approaches? How might educators decide what is most appropriate?*

> Embedded versions of the universal tools, designated supports, and accommodations are provided digitally through the test delivery system while non-embedded versions are provided at the local level through means other than the test delivery system. The choice between embedded and non-embedded universal tools and designated supports should be based on the individual student's needs. The decision should reflect the student's prior use of, and experience with, both embedded and non-embedded universal tools, designated supports, and accommodations. It is important to note that although Print on Demand is a non-embedded accommodation, permission for students to request printing must first be set in Test Information Distribution Engine (TIDE) or the state's comparable platform

4. *Who determines how non-embedded accommodations (such as read aloud) are provided?*

> IEP teams and educators make decisions about non-embedded accommodations. These teams (or educators for 504 plans) provide evidence of the need for accommodations and ensure that they are noted on the IEP or 504 plan (see *Guidelines,* pages 15-17). States are responsible for ensuring that districts and schools follow Smarter Balanced guidance on the implementation of these accommodations (see [professional development materials]).

5. *Are any students eligible to use text-to-speech for ELA reading passages on the Smarter Balanced assessments?*

> For students in grades 3-5, text-to-speech and read-aloud are not available on ELA reading passages. The use of text-to-speech (or read aloud) on ELA reading passages for grades 3-5 will result in invalid scores. In grades 6-8 and 11, text-to-speech and read-aloud are available

for ELA reading passages as an accommodation for students whose need is documented on an IEP or 504 plan (see *Guidelines*, pages 10 and 15), subject to each member state's laws, regulations, and policies. Text-to-speech and read-aloud for ELA reading passages is not available for ELLs (unless the student has an IEP or 504 plan). Whenever text-to-speech is used, appropriate headphones must be available to the student, unless the student is tested individually in a separate setting.

6. *Why are some accommodations that were previously allowed for my state assessment not listed in the Smarter Balanced Usability, Accessibility, and Accommodations Guidelines?*

    After examining the latest research and conducting numerous discussions with external and state experts, Smarter Balanced member states approved a list of universal tools, designated supports, and accommodations applicable to the current design and constructs being measured by its tests and items within them. Upon review of new research findings or other evidence applicable to accessibility and accommodations considerations, the list of specific universal tools, designated supports, and accommodations approved by Smarter Balanced may be subject to change. The Consortium will establish a standing committee, including members from Governing States, to review suggested adjustments to the list of universal tools, designated supports, and accommodations to determine whether changes are warranted.

    Proposed changes to the list of universal tools, designated supports, and accommodations will be brought to Governing States for review, feedback, and approval. Furthermore, states may issue temporary approvals (i.e., one summative assessment administration) for unique accommodations for individual students.

    State leads will evaluate formal requests for unique accommodations and determine whether the request poses a threat to the measurement of the construct. The formal requests will include documentation of the student need, the specific nature of the universal tools, designated supports, or accommodations, and the plan for follow-up monitoring of use. Upon issuing a temporary approval, the State will send documentation of the approval to the Consortium. The Consortium will consider all state-approved temporary accommodations as part of the Consortium's accommodations review process. The Consortium will provide to member states a list of the temporary accommodations issued by states that are not Consortium-approved accommodations.  In subsequent years, states will not be able to offer as a temporary accommodation any temporary accommodation that has been rejected by the Consortium.

7. *Under which conditions may a state elect not to make available to its students an accommodation that is allowed by Smarter Balanced?*

    The Consortium recognizes that there should be a careful balance between the need for uniformity among member states and the need for states to maintain their autonomy. To maintain this balance, individual states may elect not to make available an accommodation that is in conflict with the member state's laws, regulations, or policies.

8. *Can states allow additional universal tools, designated supports, or accommodations to individual students on a case by case basis?*

    Yes, only in certain restricted and emergent circumstances. To address emergent issues that arise at the local level, authorized staff in member states will have the authority to approve temporary unique testing conditions for individual students. Because it is unknown whether a temporarily provided universal tool, designated support or accommodation actually belongs

in the defined categories, all such temporary testing conditions are considered to be unique accommodations. Authorized state staff includes only those individuals who are familiar with the constructs the Smarter Balanced assessments are measuring, so that students are not inadvertently provided with universal tools, designated supports, or accommodations that violate the constructs being measured.

The unique accommodations approved by a state for individual students will be submitted to Smarter Balanced for review. Temporary unique accommodations accepted by Smarter Balanced will be incorporated into the official guidelines released by Smarter Balanced in the following year. Authorized state staff members are not to add any universal tools, designated supports, or accommodations to the Smarter Balanced *Guidelines*; only the Smarter Balanced Consortium may do so.

9. *What is to be done for special cases of "sudden" physical disability?*

   One exception to the IEP or 504 requirement is for students who have had a physical injury (e.g., broken hand or arm) that impairs their ability to use a computer. For these situations, students may use the speech-to-text or scribe accommodations (if deemed appropriate based on the student having had sufficient experience with the use of the accommodations) (see *Guidelines*, page 13).

10. *Who reviewed the Smarter Balanced Guidelines?*

    In addition to individuals and officials from the Smarter Balanced governing states, several organizations and their individual members provided written feedback on the guidelines:

    - American Federation of Teachers
    - California School for the Blind
    - California School for the Deaf
    - Californians Together
    - California State Teach
    - Center for Applied Special Technology
    - Center for Law and Education
    - Conference of Educational Administrators of Schools and Programs for the Deaf
    - Council for Exceptional Children
    - Council of the Great City Schools
    - Council of Parent Attorneys and Advocates
    - Learning Disabilities Association of Maryland
    - Mexican American Legal Defense and Education Fund
    - Missouri School Boards' Association
    - Missouri Council of Administrators of Special Education
    - National Center for Learning Disabilities
    - The Advocacy Institute
    - The National Hispanic University

11. *Where can a person go to get more information about making decisions on the use of designated supports and accommodations?*

    Practice tests provide students with experiences that are critical for success in navigating the platform easily. The practice tests may be particularly important for those students who will be using designated supports or accommodations, because the practice tests can provide data that may be useful in determining whether a student might benefit from the use of a

particular designated support or accommodation. Smarter Balanced practice tests are available at http://www.smarterbalanced.org/pilot-test/.

In addition, it is recommended that decision makers refer to professional development materials provided by Smarter Balanced or state offices on the *Individual Student Assessment Accessibility Profile (ISAAP)* or state-developed process, as well as other state-developed materials consistent with the Smarter Balanced *Implementation Guide.*

Additional information on the decision-making process, and ways to promote a thoughtful process rather than an automatic reliance on a checklist or menu, is available through materials developed by groups of states.[22]

12. *What security measures need to be taken before, during, and after the assessment for students who use universal tools, designated supports, or accommodations?*

Test security involves maintaining the confidentiality of test questions and answers, and is critical in ensuring the integrity of a test and validity of test results. Ensuring that only authorized personnel have access to the test and that test materials are kept confidential is critical in technology-based assessments. In addition, it is important to guarantee that (a) students are seated in such a manner that they cannot see each other's terminals, (b) students are not able to access any unauthorized programs or the Internet while they are taking the assessment, and (c) students are not able to access any externally-saved data or computer shortcuts while taking the test. Prior to testing, the IEP team should check on compatibility of assistive technology devices and make appropriate adjustments if necessary. When a non-embedded designated support or accommodation is used that involves a human having access to items (e.g., reader, scribe), procedures must be in place to ensure that the individual understands and has agreed to security and confidentiality requirements. Test administrators need to (a) keep testing materials in a secure place to prevent unauthorized access, and (b) keep all test content confidential and refrain from sharing information or revealing test content.

Printed test items/stimuli, including embossed Braille printouts, must be collected and inventoried at the end of each test session and securely shredded immediately. DO NOT keep printed test items/stimuli for future test sessions.

The following test materials must be securely shredded immediately after each testing session and may not be retained from one testing session to the next:

- Scratch paper and all other paper handouts written on by students during testing;

---

[22] These materials were developed by collaboratives of states to address decision making for students with disabilities, ELLs, and ELLs with disabilities:

- *Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of Students with Disabilities* (3rd ed.). Washington, DC: Assessing Special Education Students State Collaborative on Assessment and Student Standards, Council of Chief State School Officers. Available at: www.ccsso.org/Resources/Programs/Assessing_Special_Education_Students_(ASES).html.
- *Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of English Language Learners.* Washington, DC: Washington, DC: Assessing English Language Learners State Collaborative on Assessment and Student Standards, Council of Chief State School Officers. Available at: www.ccsso.org?Resources?Programs?English_Language_Learners_(ELL).html.
- *Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of English Language Learners with Disabilities.* Washington, DC: Assessing Special Education Students and English Language Learners State Collaboratives on Assessment and Student Standards, Council of Chief State School Officers. Available at www.ccsso.org/Resources/Publications/Accommodations_Manual_How_to_Select_Administer_and_Evaluate_Use_of _Accommodations_for_Instruction_and_Assessment_of_English_Language_Learners_with_Disabilities.html.

— Please note, for mathematics and ELA performance tasks, if a student needs to take the performance task in more than one session, scratch paper may be collected at the end of each session, securely stored, and made available to the student at the next performance task testing session. Once the student completes the performance task, the scratch paper must be collected and securely destroyed to maintain test security.

- Any reports or other documents that contain personally identifiable student information;
- Printed test items or stimuli.

Additional information on this topic is provided in the Test Administration Manual (TAM).

13. *Who is supposed to input information about designated supports and accommodations into the Test Information Distribution Engine (TIDE) or into a state's comparable platform? How is the information verified?*

Generally a school or district will designate a person to enter information into the TIDE or the state's comparable platform. Often this person is a test coordinator. For those students for whom an IEP team (or educator developing the 504 plan) is identifying designated supports as well as accommodations, that team or educator is responsible for ensuring that information from the IEP (or 504 plan) is entered appropriately so that all embedded accommodations can be activated prior to testing.

Entry of information for IEP and 504 students can be accomplished by identifying one person from the team to enter information or by providing information to the person designated by the school or district to enter data into the TIDE. For students who are ELLs, an educator who knows the student well and is familiar with the instructional supports used in the classroom should provide information to the person designated to enter information into the TIDE.

14. *Are there any supplies that schools need to provide so that universal tools, designated supports, and accommodations can be appropriately implemented?*

Schools should determine the number of headphones they will provide (for text-to-speech, as well as for the listening test) and other non-embedded universal tools (e.g., thesaurus), designated supports (e.g., bilingual dictionary), and accommodations (e.g., multiplication table) for students. An alternative is to identify these as items that students will provide on their own.

15. *What happens when accommodations listed in the Usability, Accessibility, and Accommodations Guidelines do not match any accommodations presented in the student's IEP?*

IEP teams should consider accommodations a student needs in light of the Smarter Balanced Guidelines. If it is decided that a specific accommodation is needed that is not included in the Guidelines, the team should submit a request to the state. The state contact will judge whether the proposed accommodation poses a threat to the constructs measured by the Smarter Balanced assessments; based on that judgment the state contact will either issue a temporary approval or will deny the request. Temporary approvals will be forwarded to a standing committee; this committee makes a recommendation to the Governing States about future incorporation of new accommodations into the Smarter Balanced Guidelines.

## Universal Tools and Designated Supports FAQs (Available to All Students)

16. *Is the digital notepad universal tool fully available for ELA and Math? Will a student's notes be saved if the student takes a 20-minute break?*

    The digital notepad is available on all items across both content areas. As long as a student or test administrator activates the test within the 20-minute break window, the notes will still be there. There is no limit on the number of pauses that a student can take in one test sitting.

17. *For the global notes universal tool, if a student takes a break of 20 minutes do the notes disappear?*

    Global notes, which are used for ELA performance tasks only, will always be available until the student submits the test, regardless of how long a break lasts or how many breaks are taken.

18. *For the highlighter universal tool, if a student pauses a test for 20-minutes, do the highlighter marks disappear?*

    If a student is working on a passage or stimulus on a screen and pauses the test for 20 minutes to take a break, the student will still have access to the information visible on that particular screen. However, students do lose access to any information highlighted on a previous screen.

19. *How are students made aware that the spell check universal tool (for ELA) and the math universal tools (i.e., calculator) are available when moving from item to item?*

    When appropriate, math items include universal tools available for students to use. For the spell check tool, a line will appear under misspelled words.

20. *For the zoom universal tool, is the default size specific to certain devices? Will the test administrator's manual provide directions on how to do this adjustment?*

    The default size is available to all students and is not specific to certain devices. Information on how to use the zoom universal tool is included in the directions at the beginning of each test. Please note that in addition to zoom, students may have access to magnification, which is a non-embedded designated support.

21. *For the English glossary universal tool, how are terms with grade- and context-appropriate definitions made evident to the student?*

    Selected terms have a light rectangle around them. If a student hovers over the terms, the terms with the attached glossary are highlighted. A student can click on the terms and a pop-up window will appear. In addition, a student can click on the audio button next to each term to hear it.

22. *For the mark-for-review universal tool, will selections remain visible after a 20-minute break?*

    If a student takes a break for longer than 20 minutes, the student will not be able to access items from previous screens.

23. *Can universal tools be turned off if it is determined that they will interfere with the student's performance on the assessment?*

> Yes. If an adult (or team) determines that a universal tool might be distracting or that students do not need to or are unable to use them. This information must be noted in TIDE prior to test administration.

## FAQs Pertaining to English Language Learners (ELLs)

24. *How are the language access needs of ELLs addressed in the Smarter Balanced Usability, Accessibility, and Accommodations Guidelines?*

> The language access needs of ELLs are addressed through the provision of numerous universal tools and designated supports. These include universal tools such as English dictionaries for full writes and English glossaries, and designated supports such as translated test directions and glossaries. These are not considered accommodations in the Smarter Balanced assessment system. No accommodations are available for ELLs on the Smarter Balanced assessments; accommodations are only available to students with disabilities and ELLs with disabilities.

25. *Is text-to-speech available for ELLs to use?*

> Text-to-speech is available as a designated support to all students (including ELLs) for whom an adult or team has indicated it is needed for math items and for ELA items (but not ELA reading passages). Text-to-speech for ELA reading passages is available for an ELL in grades 6-8 or 11 only if the student has an IEP or 504 plan. For text-to-speech to be available for an ELL, it must be entered into the TIDE.

26. *What languages are available to ELLs in text-to-speech?*

> Text-to-speech is currently available only in English. However, the translated glossaries include an audio component automatically available to any student with the translated glossaries embedded designated support.

27. *For which content areas will the Consortium provide translation supports for students whose primary language is not English?*

> For Mathematics, the Consortium will provide full translations in American Sign Language, stacked translations in Spanish (with the Spanish translation presented directly above the English item), and primary language pop-up glossaries in various languages and dialects including Spanish, Vietnamese, Arabic, Tagalog, Ilokano, Cantonese, Mandarin, Korean, Punjabi, Russian, and Ukrainian. For the Listening portion of the English Language Arts assessment, Smarter Balanced will provide full translations in American Sign Language delivered digitally through the test delivery system.

> Only translations that have gone through the translation process outlined in the Smarter Balanced Translation framework would be an accepted support (http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/Translation-Accommodations-Framework-for-Testing-ELL-Math.pdf).

*28. Does a student need to be identified as an English language learner in order to receive translation and language supports? What about foreign language exchange students?*

Translations and language supports are provided as universal tools and designated supports. Universal tools are available to all students. Designated supports are available to those students for whom an adult (or team) has determined a need for the support. Thus, these are available to all students, regardless of their status as an ELL. Foreign language exchange students would have access to all universal tools and those designated supports that have been indicated by an adult (or team).

*29. For the translated test directions designated support, what options are available for students who do not understand the language available in the digital format? Can a human reader of directions in the native language be provided?*

If a student needs a read aloud/text-to-speech accommodation in another language, then the test directions should be provided in that other language. The reader or text-to-speech device must be able to provide the directions in the student's language without difficulty due to accent or register. To ensure quality and standardized directions, the reader or text-to-speech device should only use directions that have undergone professional translation by the Consortium prior to testing. Smarter Balanced is providing a PDF of the translated test directions in each of the languages supported by the translated glossary designated support: Spanish, Vietnamese, Arabic, Tagalog, Ilokano, Cantonese, Mandarin, Korean, Punjabi, Russian, and Ukrainian.

*30. How is the translations glossary non-embedded designated support different from the bilingual dictionary?*

The translations glossary non-embedded designated support includes the customized translation of pre-determined construct-irrelevant terms that are most challenging to English language learners. The translation of the terms is context-specific and grade-appropriate. Bilingual dictionaries often do not provide context-specific information nor are they customized. In addition, the translated glossary includes an audio support.

*31. Will translations be available in language dialects/variants?*

Translated glossaries will be available in different languages and dialects including Spanish, Vietnamese, Arabic, Tagalog, Ilokano, Cantonese, Mandarin, Korean, Punjabi, Russian, and Ukrainian.

## FAQs Pertaining to Students with Disabilities

*32. What accommodations are available for students with disabilities (including ELLs with disabilities)?*

Students with disabilities (including those who are ELLs) can use embedded accommodations (e.g., American Sign Language, braille, speech-to-text) and non-embedded accommodations (e.g., abacus, alternate response options) that have been documented on an IEP or 504 accommodations plan. These students also may use universal tools and designated supports. A full list of accommodations can be found in the Guidelines documents, tables 5 and 6.

*33. Is an embedded ASL accommodation available on ELA items that are not part of the Listening test?*

The embedded ASL accommodation is not currently available on any ELA items that are not part of the Listening claim. For the Listening test, a deaf or hard of hearing student who has a documented need in an IEP or 504 plan may use ASL.

*34. Will sign languages other than ASL (including signing in other languages) be available?*

Currently, only ASL is available.

*35. Can interpreters be used for students who are deaf or hard of hearing who do not use ASL?*

Smarter Balanced has consulted with external experts who have unanimously advised against this practice. Research indicates severe challenges with standardization and quality.

*36. What options do districts have for administering Smarter Balanced assessments to students who are blind?*

Students who are blind and who prefer to use braille should have access to either refreshable braille (only for ELA) or embosser-created braille (for ELA or math). For those students who are blind and prefer to use text-to-speech, access to text-to-speech should be provided for the math test, and for ELA items only (text-to-speech is not permitted on ELA reading passages without a specific documented need in the student's IEP or 504 plan). Text-to-speech use for ELA reading passages is only permitted for those students in grades 6-8 and 11. Students should participate in the decision about the accommodation they prefer to use, and should be allowed to change during the assessment if they ask to do so. Students can have access to both Braille and text-to-speech that is embedded in the Smarter Balanced assessment system.

*37. Why is the non-embedded abacus an accommodation for the non-calculator items? Doesn't an abacus serve the same function as a calculator?*

An abacus is similar to the sighted student using paper and pencil to write a problem and do calculations. The student using the abacus has to have an understanding of number sense and must know how to do calculations with an abacus.

*38. Can students without documented disabilities who have had a sudden injury use any of the Smarter Balanced accommodations?*

Students without documented disabilities who have experienced a physical injury that impairs their ability to use a computer may use some accommodations, provided they have had sufficient experience with them. Both speech-to-text and scribe are accommodations that are available to students who have experienced a physical injury such as a broken hand or arm, or students who have become blind through an injury and have not had sufficient time to learn braille. Prior to testing a student with a sudden physical injury, regardless of whether a 504 plan is started, Test Administrators should contact their district test coordinator or other authorized individuals to ensure the test registration system accurately describes the student's status and any accommodations that the student requires.

*39. How will the test administrator know prior to testing that the print on demand accommodation may be needed?*

The test administrator will know this information prior to testing because accommodations need to be documented beforehand and print on demand is an accommodation. Any accommodations – including both embedded and non-embedded accommodations – need to be entered into the TIDE. The print on demand accommodation applies to either passages/stimuli or items, or both.

*40. For the print on demand accommodation, how are student responses recorded – by a teacher using a computer or some other method?*

The method of recording student responses depends on documentation in the IEP or 504 plan (e.g., after first recording responses on the paper version, the student could enter responses into the computer or the teacher could enter responses into the computer.) Anyone who is designated to enter responses into the computer must have read, agreed to, and signed a test security agreement.

*41. How do state officials monitor training and qualifications for the non-embedded read aloud accommodation?*

States will need to develop processes and procedures to monitor training and the qualifications of individuals who provide the read aloud accommodation when text-to-speech is not appropriate for a student. State officials can use the Smarter Balanced audio guidelines available online to obtain additional information about recommended processes to follow (http://www.smarterbalanced.org/smarter-balanced-assessments/#item).

*42. If students are using their own devices that incorporate word prediction, will this impact their score?*

The students' score will not be affected under these circumstances. Students using these devices must still use their knowledge and skills to review and edit their answers.

*43. How are assistive technology (AT) devices certified for use for the Smarter Balanced assessments?*

Assistive technology device manufacturers may use the Smarter Balanced practice test as a method of determining if a device works with the assessment. In addition, schools and districts can use the practice test to evaluate devices to ensure their functions are consistent with those allowed in the UAAG.

# Revision Log

Updates to the *Smarter Balanced Usability, Accessibility, and Accommodations Guidelines* are captured in this Revision Log. Updates are based on requests from states that do not impact policy. Any changes impacting policy require discussion and vote by Governing States. Updates captured in the Revision Log are separated into two categories:

- **Clarification:** Updates of this type add details to existing information included in the *Guidelines*.
- **Increased Flexibility:** Updates of this type reflect explicatory information included in the *Guidelines* that result in augmented access to Smarter Balanced assessments.

Revisions are captured in tracking tables according to category. In cases where both **Clarification** and **Increased Flexibility** edits are made, changes to the *Guidelines* will be captured in the **Increased Flexibility** tracking table.

| Section | Page | Clarification: Description of Changes | Date | Version |
|---------|------|----------------------------------------|------|---------|
| Table 3 | 9 | Consistently used the term "ELA reading passages" instead of "ELA passages" to clarify availability of text-to-speech as an embedded designated support. | 03/12/14 | 1.2 |
| Table 4 | 11 | Consistently used the term "ELA reading passages" instead of "ELA passages" to clarify availability of read aloud as a non-embedded designated support. | 03/12/14 | 1.2 |
| Table 5 | 15 | Consistently used the term "ELA reading passages" instead of "ELA passages" to clarify availability of text-to-speech as an embedded accommodation. | 03/12/14 | 1.2 |
| Table 6 | 16 | Consistently used the term "ELA reading passages" instead of "ELA passages" to clarify availability of read aloud as a non-embedded accommodation. | 03/12/14 | 1.2 |
| Table 3 | 10 | Added verbiage clarifying the audio component of translated glossaries. | 08/01/14 | 2.1 |

| Section | Page | Increased Flexibility: Description of Changes | Date | Version |
|---------|------|----------------------------------------------|------|---------|
| Table 2 | 8 | Scratch paper, the non-embedded universal tool, description has additional details regarding the performance task testing sessions: "For mathematics and ELA performance tasks, if a student needs to take the performance task in more than one session, scratch paper may be collected at the end of each session, securely stored, and made available to the student at the next performance task testing session. Once the student completes the performance task, the | 03/12/14 | 1.2 |

| Section | Page | Increased Flexibility: Description of Changes | Date | Version |
|---|---|---|---|---|
| | | scratch paper must be collected and securely destroyed to maintain test security." | | |
| Table 4 | 13 | Added information regarding the availability of translated test directions in PDF format. New accessibility resource also added to Figure 1 and Appendix A. | 08/01/14 | 2.1 |
| Table 4 | 13 | To separate setting, added that, "A specific adult, trained in a manner consistent with the TAM, can act as test proctor (test administrator) when student requires it." | 08/01/14 | 2.1 |
| Table 6 | 17 | Added information regarding the availability of noise buffers. New accessibility resource also added to Figure 1 and Appendix A. | 08/01/14 | 2.1 |
| Appendix | 24 | Added the FAQs section. | 08/01/14 | 2.1 |

Appendix H –     Small Scale Trials Technical Report

# Smarter Balanced Assessment Consortium:
## Small Scale Trials Technical Report

## Executive Summary

The Smarter Balanced Assessment Consortium seeks to develop a testing framework that assesses student performance with authentic instruments that closely resemble the classroom learning experience. This will be accomplished through the integration of technology to achieve better measures of deeper learning outcomes that have been difficult to efficiently measure in the past. The Consortium seeks to understand and extend the existing state of the art in automated scoring.

These objectives will be met with a financially sustainable model in which substantial parts of the test are automatically scored, either exclusively or in conjunction with some human process. The feedback to students is envisioned to be provided immediately after completion of the assessment. The Consortium recognizes that this quick response can be accomplished only through the use of computer-based testing and the development of automated scoring models for constructed-response items.

The Small Scale Trial represents the second study on automated scoring models.  An initial report (The Initial Analysis of the Essay Scoring Engine) examined the application of an essay scoring engine applied to four Reading items presently in use in two Consortium states.  These items were not intended to be scored using automated scoring models, but the analysis provided an initial look at items presently in use in Consortium states. Automated scoring models built using this data helped inform the models built for the fall 2012 Small Scale Trials.  The Small Scale Trial data was used to further evaluate and improve the essay scoring engine and provide a first look at the application of the propositional scoring model to Consortium items.  The results from these analyses will further inform the essay scoring model vision for the Consortium prior to the 2014 field test.

The Small Scale Trial sought to expand on the initial model building effort in a number of ways:

1. Items included in the Small Scale Trial test forms were built to Smarter Balanced specifications.
2. Items were examined in three content areas (reading, writing, and mathematics), at three grade levels (4, 7, and 11). A refined essay scoring model was applied to writing essays using multiple rubrics.
3. The propositional model was applied to short content-based constructed responses.

The purpose of this document is to present the findings from the Small Scale Trial study.


## Background

For standardized tests, we expect scores to be comparable over time and over different administrations and forms of a test. One of the main concerns of constructed-response scoring is scoring consistency. The constructed-response scoring process often includes elaborate systems for monitoring the consistency and accuracy of the scores (e.g., back-reading, validity scoring, double scoring). Even with highly structured training and monitoring, the scoring process still leaves room for differences of opinion among raters.

Computer-automated scoring (CAS) has the potential to make the scoring of constructed-response items more objective.  It also has the benefit of making constructed-response test items practical for

use in situations where scoring by human scorers is not realistic.  This is especially true for computer-adaptive testing where computer-automated scoring can immediately provide data for use in the selection of subsequent items.

Williamson, Bejar, and Hone (1999) listed a number of advantages of modern CAS systems over human scoring. With CAS systems, a given response will always receive the same results (reproducibility); the same scoring criteria are consistently applied to all responses (consistency); specific reasons and processes behind computer scoring can be traced, investigated, and manipulated (tractability); items can be constructed in a more precise fashion (item specification); responses can be evaluated at a higher level of precision and specificity (granularity); scoring criteria are better articulated, and much of the subjectivity in human scoring can be removed (objectivity); scoring outcomes are likely to be more reliable (reliability); and the scoring process can be less demanding in terms of time, resources, and cost (efficiency).

The validity of computer-assigned scores using various scoring engines has been evaluated by comparing computer-assigned scores with human scores (Attali, Powers, Freedman, Harrison, & Obetz, 2008; Bennett, Steffen, Singley, Morley, & Jacquemin, 1997; Klein, 2008; Yang, Buckendahl, Juszkiewicz, & Bhola, 2002). The quantitative methods included agreements and correlations between raters and between rater and machine scores. Scores were also compared across subgroups. Although the results varied across subjects and item types, the overall findings demonstrated that computer-assigned scores were very similar to human scores and suggested that machine scoring could facilitate the use of constructed-response items in large-scale testing programs by providing a fast, accurate, and efficient way to score responses. Attali et al. (2008) evaluated the quality of computer-automated scoring for open-ended items (those requiring a short answer of one to three sentences) of GRE® Subject Test items in biology and psychology using the c-rater™ scoring engine. The kappa agreements (agreement beyond chance) were higher for psychology questions than for biology questions; however, both the human-human agreement and human-computer agreement were moderate to high.

### Forms Design

The scope of the Small Scale Trial targeted computer-automated item scoring in grades 4, 7, and 11. Within each grade, a separate 15-item test consisting of selected-response (SR) items was constructed for reading, writing, and mathematics. Three constructed-response (CR) items (intended for machine-scoring) were included in the test forms in reading and math at each grade level. For writing, three pairs of constructed-response sets were constructed for each grade. Two sets included a brief writing CR and an essay (long) writing prompt. One set included a research CR and essay (long) writing prompt.

Table 1 shows the number of items included in the Smarter Balanced Small Scale Trials administration at each grade. Note that the same SR set was used for each writing form in a given grade.

Table 1. Number of Items Administered at Each Grade for the Smarter Balanced Small Scale Trials

| Form | Subject | Grade | Selected-Response Items | Constructed-Response Items | Total Items |
|------|---------|-------|-------------------------|----------------------------|-------------|
| 1 | Reading | 4 | 15 | 3 | 18 |
| 2 | Reading | 7 | 15 | 3 | 18 |
| 3 | Reading | 11 | 15 | 3 | 18 |
| 4 | Writing A | 4 | 15 | 2 | 17 |
| 5 | Writing B | 4 | - | 2 | 2 |
| 6 | Writing C | 4 | - | 2 | 2 |
| 7 | Writing A | 7 | 15 | 2 | 17 |
| 8 | Writing B | 7 | - | 2 | 2 |
| 9 | Writing C | 7 | - | 2 | 2 |
| 10 | Writing A | 11 | 15 | 2 | 17 |
| 11 | Writing B | 11 | - | 2 | 2 |
| 12 | Writing C | 11 | - | 2 | 2 |
| 13 | Math | 4 | 15 | 3 | 18 |
| 14 | Math | 7 | 15 | 3 | 18 |
| 15 | Math | 11 | 15 | 3 | 18 |
| | | Total | 135 | 36 | 171 |

## Data

The goal was to obtain 1000 responses for each of the five test forms in grades 4, 7, and 11. To obtain this sample, 911 schools were selected to yield a projected sample of approximately 730 schools in 23 Consortium states. In the end, 427 schools from 21 states participated. The full Sampling Plan can be found in Appendix A.

## Human Scoring

All student constructed responses were scored by two trained human scorers (100% double reads). Responses that received nonadjacent scores (e.g., a 2 and a 4) were routed to an expert scorer for a third independent reading. Since 0–1 point items could not have nonadjacent scores, each 0–1 point response was scored by two independent readers, and if the scores were not an exact match, a third expert reader scored the response.

The range finding/rubric validation responses formed the basis of the training materials. Range finding/rubric validation responses were supplemented with live responses where necessary. These responses were selected based on the way that the committees applied the rubrics to the items. All training materials were reviewed and approved by Smarter Balanced representatives or by range finding committee members.  Table 2 details the training materials developed for each item.

Table 2. Description of Training Materials

| Score Point Range | Anchor Sets (for each item) | Training Sets (for each item) | Qualifying Sets (for 1 item per grade and item type) | Qualifying Rate (exact agreement) | Validity Responses (for each item) |
|---|---|---|---|---|---|
| 0–1 | 3 responses/ score point | 1 set of 5 responses and 1 set of 10 responses (15 total responses) | 2 sets of 10 responses (20 total responses) | 90% | 10 |
| 0–2 | 3 responses/ score point | 2 sets of 10 responses (20 total responses) | 2 sets of 10 responses (20 total responses) | 80% | 10 |
| Long Writing Item | 3 responses/ score point (see below for details) | 3 single-trait sets of 5 responses 1 multi-trait set of 5 responses 1 multi-trait set of 10 responses | 4 sets of 10 responses (40 total responses) | 70% in each trait | 10 |

Each item was scored by one team of approximately nine to twelve scorers. The only exception was the long writing items, which were each scored by two teams of scorers per grade.

For each item type, scorers were required to qualify on one item per grade. For example, there were two grade 4 brief writing items. One of these two items had two qualifying sets. The scorers started with the item that had qualifying sets. Since the brief writing items are all scored on a 0–2 point scale, the scorers were required to correctly score 80% of the responses in one of the two qualifying sets. All of the scorers were able to successfully demonstrate the required level of accuracy while qualifying.

Each long writing response received scores in three traits, so some of the training material focused on one trait, while other materials considered all three traits:

- Single-Trait Long Write Anchor Sets
  - Focus and Organization (1–4 point scale): One anchor set with a minimum of three responses per score point for each item
  - Elaboration (1–4 point scale): One anchor set with a minimum of three responses per score point for each item
  - Conventions (1–2 point scale): One anchor set with a minimum of three responses per score point per grade
- Single-Trait Training Sets (Scorers scored one trait for these sets.)
  - Focus and Organization: One 5-response training set per item
  - Elaboration: One 5-response training set per item
  - Conventions: One 5-response  training set per grade
- Multi-Trait Training Sets (Scorers scored all three traits for these sets.)
  - Set 1: One 5-response training set per item
  - Set 2: One 10-response training set per item
- Multi-Trait Qualifying Sets (Scorers scored all three traits for these sets.)
  - Scorers qualified on one item per grade. The vendors and Smarter Balanced determined which item during range finding.
  - Four 10-response qualifying sets per item
  - In order to quality, scorers had to demonstrate sufficient accuracy as follows:
    - 70% exact agreement in Focus and Organization on one of the four sets
    - 70% exact agreement  in Elaboration on one of the four sets
    - 70% exact agreement  in Conventions on one of the four sets

Throughout scoring, several measures were taken to evaluate and monitor quality control. Scorers were given ongoing feedback and retraining based on the quality control measures. These quality control measures included:

- **Scoring Summary Report:** Daily and cumulative reports provided inter-rater agreement rates and score point distributions by scorer and room. The cumulative results are reported in Table 3.

- **Team Leader Read-behinds:** Team Leaders and Scoring Directors "spot checked" the scorers' performance by reviewing approximately 10% of the responses read by each scorer. If the supervisor disagreed with the score given by the scorer, the supervisor corrected the score and, as appropriate, shared the response and the corrected score with the scorer as an opportunity to provide ongoing feedback and improve scoring accuracy.

- **Validity:** Validity responses were pre-scored based on the way that the range finding committees applied the rubrics. They were distributed to the scorers throughout the scoring window, although they were front-loaded to help ensure that every scorer received every validity response. The responses were randomly selected; as a result, some validity sets did not contain all possible score points. Scorers were not able to distinguish validity responses from live student responses, making this a powerful measure of quality control. At least ten validity responses were implemented for each item. The validity results are reported in Table 4.

Table 3. Human Scoring Item Summary Report

## Item Summary Report
## Smarter Balanced Assessment Consortium
## Small Scale Trials

| Subject | Grade | Domain | ID | Inter-Rater Reliability | | | | Score Point Distribution | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 2X | %EX | %AD | %NA | Total | %0 | %1 | %2 | %3 | %4 | %B[1] | %F[2] | %M[3] | %N[4] | %T[5] | %U[6] |
| Brief Write | Grade 04 | | 43403 | 2,134 | 81 | 18 | 0 | 2,134 | 15 | 46 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Brief Write | Grade 07 | | 43497 | 1,412 | 78 | 22 | 1 | 1,412 | 50 | 33 | 10 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 0 |
| Brief Write | Grade 07 | | 43964 | 1,424 | 85 | 15 | 0 | 1,424 | 8 | 44 | 43 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 0 |
| Brief Write | Grade 11 | | 43446 | 1,166 | 84 | 16 | 0 | 1,166 | 8 | 50 | 41 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Long Write | Grade 04 | Organization | 43504 | 2,088 | 88 | 12 | 0 | 2,088 | 0 | 46 | 32 | 5 | 0 | 0 | 0 | 0 | 4 | 12 | 1 |
| | | Conventions | | 2,088 | 81 | 19 | 0 | 2,088 | 24 | 33 | 26 | 0 | 0 | 0 | 0 | 0 | 4 | 12 | 1 |
| | | Elaboration | | 2,088 | 88 | 11 | 0 | 2,088 | 0 | 45 | 33 | 4 | 0 | 0 | 0 | 0 | 4 | 12 | 1 |
| Long Write | Grade 04 | Organization | 43334 | 2,114 | 86 | 14 | 0 | 2,114 | 0 | 39 | 40 | 9 | 0 | 0 | 0 | 0 | 2 | 9 | 0 |
| | | Conventions | | 2,114 | 81 | 18 | 0 | 2,114 | 30 | 35 | 23 | 0 | 0 | 0 | 0 | 0 | 2 | 9 | 0 |
| | | Elaboration | | 2,114 | 86 | 14 | 0 | 2,114 | 0 | 40 | 40 | 9 | 0 | 0 | 0 | 0 | 2 | 9 | 0 |
| Long Write | Grade 04 | Organization | 43284 | 2,084 | 85 | 14 | 0 | 2,084 | 0 | 39 | 26 | 10 | 2 | 0 | 0 | 0 | 4 | 19 | 1 |
| | | Conventions | | 2,084 | 85 | 14 | 0 | 2,084 | 29 | 27 | 19 | 0 | 0 | 0 | 0 | 0 | 4 | 19 | 1 |
| | | Elaboration | | 2,084 | 86 | 14 | 0 | 2,084 | 0 | 38 | 26 | 10 | 1 | 0 | 0 | 0 | 4 | 19 | 1 |
| Long Write | Grade 07 | Organization | 43438 | 1,396 | 83 | 16 | 0 | 1,396 | 0 | 46 | 34 | 6 | 0 | 0 | 0 | 0 | 1 | 11 | 1 |
| | | Conventions | | 1,396 | 77 | 22 | 0 | 1,396 | 21 | 34 | 32 | 0 | 0 | 0 | 0 | 0 | 1 | 11 | 1 |
| | | Elaboration | | 1,396 | 84 | 15 | 0 | 1,396 | 0 | 47 | 34 | 6 | 0 | 0 | 0 | 0 | 1 | 11 | 1 |

[1] B indicates a condition code designating this response is blank.

[2] F is a condition code indicating a response is not in English.

[3] M indicates that a response is off purpose. Please note that this condition code is only applicable to long writing items. If a long write receives a code of M, it will still be scored for conventions. If a long write receives any other non-scorable code, it will not be scored in any domain.

[4] N indicates the response was non-scorable for any reason. For example, this code would be appropriate if a student copied text.

[5] T is used when a response is off topic. For example, a student responds that they hate pizza, when the item was about helicopters.

[6] U indicates that the response is unintelligible. This condition code would be appropriate when a student submits random keystrokes or undecipherable text.

| Subject | Grade | Domain | ID | 2X | %EX | %AD | %NA | Total | %0 | %1 | %2 | %3 | %4 | %B[1] | %F[2] | %M[3] | %N[4] | %T[5] | %U[6] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Long Write | Grade 07 | Organization | 43703 | 1,396 | 88 | 12 | 0 | 1,396 | 0 | 53 | 33 | 2 | 0 | 0 | 0 | 0 | 1 | 10 | 0 |
| | | Conventions | | 1,396 | 79 | 20 | 1 | 1,396 | 34 | 31 | 24 | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 0 |
| | | Elaboration | | 1,396 | 88 | 12 | 0 | 1,396 | 0 | 53 | 34 | 1 | 0 | 0 | 0 | 0 | 1 | 10 | 0 |
| Long Write | Grade 07 | Organization | 43469 | 1,384 | 84 | 16 | 0 | 1,384 | 0 | 39 | 34 | 9 | 2 | 0 | 0 | 3 | 2 | 10 | 0 |
| | | Conventions | | 1,384 | 81 | 19 | 0 | 1,384 | 32 | 35 | 21 | 0 | 0 | 0 | 0 | 0 | 2 | 10 | 0 |
| | | Elaboration | | 1,384 | 83 | 17 | 0 | 1,384 | 0 | 41 | 33 | 9 | 2 | 0 | 0 | 3 | 2 | 10 | 0 |
| Long Write | Grade 11 | Organization | 43632 | 1,142 | 86 | 14 | 0 | 1,142 | 0 | 39 | 37 | 12 | 0 | 0 | 0 | 0 | 1 | 9 | 2 |
| | | Conventions | | 1,142 | 81 | 19 | 0 | 1,142 | 15 | 31 | 42 | 0 | 0 | 0 | 0 | 0 | 1 | 9 | 2 |
| | | Elaboration | | 1,142 | 86 | 14 | 0 | 1,142 | 0 | 40 | 37 | 11 | 0 | 0 | 0 | 0 | 1 | 9 | 2 |
| Long Write | Grade 11 | Organization | 43635 | 1,134 | 84 | 16 | 0 | 1,134 | 0 | 28 | 43 | 19 | 0 | 0 | 0 | 0 | 0 | 7 | 3 |
| | | Conventions | | 1,134 | 76 | 24 | 0 | 1,134 | 17 | 35 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 3 |
| | | Elaboration | | 1,134 | 84 | 16 | 0 | 1,134 | 0 | 29 | 43 | 18 | 0 | 0 | 0 | 0 | 0 | 7 | 3 |

## Item Summary Report
## Smarter Balanced Assessment Consortium
## Small Scale Trials

| | | | | Inter-Rater Reliability | | | | Score Point Distribution | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Domain | ID | 2X | %EX | %AD | %NA | Total | %0 | %1 | %2 | %3 | %4 | %B[1] | %F[2] | %M[3] | %N[4] | %T[5] | %U[6] |
| Long Write | Grade 11 | Organization | 43479 | 1,146 | 85 | 15 | 0 | 1,146 | 0 | 36 | 42 | 13 | 0 | 0 | 0 | 0 | 1 | 7 | 1 |
| | | Conventions | | 1,146 | 80 | 20 | 0 | 1,146 | 21 | 35 | 35 | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 1 |
| | | Elaboration | | 1,146 | 84 | 16 | 0 | 1,146 | 0 | 43 | 40 | 9 | 0 | 0 | 0 | 0 | 1 | 7 | 1 |
| Mathematics | Grade 04 | | 43572 | 2,126 | 95 | 5 | 0 | 2,126 | 56 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| Mathematics | Grade 04 | | 43564 | 1,438 | 100 | 0 | 0 | 1,438 | 89 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 1 |
| Mathematics | Grade 04 | | 43173 | 2,126 | 98 | 2 | 0 | 2,126 | 81 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |

[1] B indicates a condition code designating this response is blank.

[2] F is a condition code indicating a response is not in English.

[3] M indicates that a response is off purpose. Please note that this condition code is only applicable to long writing items. If a long write receives a code of M, it will still be scored for conventions. If a long writes receives any other non-scorable code, it will not be scored in any domain.

[4] N indicates the response was non-scorable for any reason. For example, this code would be appropriate if a student copied text.

[5] T is used when a response is off topic. For example, a student responds that they hate pizza, when the item was about helicopters.

[6] U indicates that the response is unintelligible. This condition code would be appropriate when a student submits random keystrokes or undecipherable text.

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mathematics | Grade 07 | 43551 | 1,430 | 92 | 8 | 0 | 1,430 | 79 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Mathematics | Grade 07 | 43555 | 1,422 | 92 | 8 | 0 | 1,422 | 62 | 34 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Mathematics | Grade 07 | 43557 | 1,422 | 97 | 3 | 0 | 1,422 | 72 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Mathematics | Grade 07 | 43639 | 1,422 | 96 | 4 | 0 | 1,422 | 63 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Mathematics | Grade 11 | 43559 | 1,164 | 99 | 1 | 0 | 1,164 | 94 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 |
| Mathematics | Grade 11 | 43552 | 1,122 | 100 | 0 | 0 | 1,122 | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 1 |
| Mathematics | Grade 11 | 43546 | 1,146 | 97 | 3 | 0 | 1,146 | 75 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 3 |
| Reading | Grade 04 | 43707 | 2,134 | 91 | 7 | 2 | 2,134 | 53 | 22 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Reading | Grade 04 | 43412 | 2,124 | 91 | 9 | 0 | 2,124 | 59 | 33 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 |
| Reading | Grade 04 | 43416 | 2,112 | 93 | 6 | 1 | 2,112 | 43 | 38 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Reading | Grade 07 | 43248 | 1,430 | 80 | 20 | 0 | 1,430 | 53 | 41 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Reading | Grade 07 | 43445 | 1,426 | 85 | 13 | 2 | 1,426 | 67 | 20 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 |
| Reading | Grade 07 | 43422 | 1,410 | 81 | 13 | 6 | 1,410 | 66 | 20 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Reading | Grade 11 | 43297 | 1,170 | 86 | 11 | 3 | 1,170 | 56 | 28 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 |
| Reading | Grade 11 | 43435 | 1,164 | 87 | 10 | 3 | 1,164 | 67 | 24 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Reading | Grade 11 | 43397 | 1,158 | 87 | 10 | 3 | 1,158 | 74 | 13 | 5 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 1 |
| Research | Grade 04 | 43280 | 2,102 | 93 | 7 | 0 | 2,102 | 70 | 18 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 |
| Research | Grade 07 | 43468 | 1,406 | 95 | 5 | 0 | 1,406 | 35 | 23 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| Research | Grade 11 | 43491 | 1,154 | 89 | 10 | 0 | 1,154 | 28 | 41 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 |

Table 4. Validity Summary Report for Human Scoring

**Validity Summary Report**
**Smarter Balanced Assessment Consortium**
**Small Scale Trials**

| Subject | Grade | Domain | ID | #R | #NA LO | #LO | #EX | #HI | #NA HI | %NA LO | %LO | %EX | %HI | %NA HI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brief Write | Grade 04 | | 43403 | 120 | 0 | 1 | 111 | 8 | 0 | 0 | 1 | 93 | 7 | 0 |
| Brief Write | Grade 07 | | 43497 | 118 | 0 | 5 | 99 | 14 | 0 | 0 | 4 | 84 | 12 | 0 |
| Brief Write | Grade 07 | | 43964 | 152 | 0 | 5 | 145 | 2 | 0 | 0 | 3 | 95 | 1 | 0 |
| Brief Write | Grade 11 | | 43446 | 131 | 0 | 8 | 120 | 3 | 0 | 0 | 6 | 92 | 2 | 0 |
| Long Write | Grade 04 | Organization | 43504 | 267 | 0 | 16 | 250 | 1 | 0 | 0 | 6 | 94 | 0 | 0 |
| | | Elaboration | 43504 | 267 | 0 | 17 | 249 | 1 | 0 | 0 | 6 | 93 | 0 | 0 |
| | | Conventions | 43504 | 267 | 0 | 27 | 222 | 18 | 0 | 0 | 10 | 83 | 7 | 0 |
| Long Write | Grade 04 | Organization | 43334 | 285 | 0 | 9 | 271 | 5 | 0 | 0 | 3 | 95 | 2 | 0 |
| | | Elaboration | 43334 | 285 | 0 | 10 | 271 | 4 | 0 | 0 | 4 | 95 | 1 | 0 |
| | | Conventions | 43334 | 285 | 0 | 4 | 238 | 42 | 1 | 0 | 1 | 84 | 15 | 0 |
| Long Write | Grade 04 | Organization | 43284 | 285 | 0 | 4 | 278 | 3 | 0 | 0 | 1 | 98 | 1 | 0 |
| | | Elaboration | 43284 | 285 | 0 | 5 | 278 | 2 | 0 | 0 | 2 | 98 | 1 | 0 |
| | | Conventions | 43284 | 285 | 0 | 9 | 270 | 6 | 0 | 0 | 3 | 95 | 2 | 0 |
| Long Write | Grade 07 | Organization | 43438 | 280 | 1 | 11 | 241 | 27 | 0 | 0 | 4 | 86 | 10 | 0 |
| | | Elaboration | 43438 | 280 | 1 | 11 | 241 | 27 | 0 | 0 | 4 | 86 | 10 | 0 |
| | | Conventions | 43438 | 280 | 1 | 11 | 232 | 35 | 1 | 0 | 4 | 83 | 13 | 0 |
| Long Write | Grade 07 | Organization | 43703 | 300 | 0 | 9 | 245 | 46 | 0 | 0 | 3 | 82 | 15 | 0 |
| | | Elaboration | 43703 | 300 | 0 | 10 | 244 | 46 | 0 | 0 | 3 | 81 | 15 | 0 |
| | | Conventions | 43703 | 300 | 0 | 19 | 262 | 19 | 0 | 0 | 6 | 87 | 6 | 0 |
| Long Write | Grade 07 | Organization | 43469 | 266 | 3 | 4 | 226 | 33 | 0 | 1 | 2 | 85 | 12 | 0 |
| | | Elaboration | 43469 | 266 | 3 | 5 | 226 | 32 | 0 | 1 | 2 | 85 | 12 | 0 |
| | | Conventions | 43469 | 266 | 3 | 8 | 234 | 21 | 0 | 1 | 3 | 88 | 8 | 0 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Long Write | Grade 11 | Organization | 43632 | 300 | 0 | 13 | 253 | 34 | 0 | 0 | 4 | 84 | 11 | 0 |
| | | Elaboration | 43632 | 300 | 0 | 18 | 263 | 19 | 0 | 0 | 6 | 88 | 6 | 0 |
| | | Conventions | 43632 | 300 | 1 | 21 | 253 | 25 | 0 | 0 | 7 | 84 | 8 | 0 |
| Long Write | Grade 11 | Organization | 43635 | 323 | 0 | 6 | 284 | 33 | 0 | 0 | 2 | 88 | 10 | 0 |
| | | Elaboration | 43635 | 323 | 0 | 10 | 284 | 29 | 0 | 0 | 3 | 88 | 9 | 0 |
| | | Conventions | 43635 | 323 | 0 | 29 | 248 | 46 | 0 | 0 | 9 | 77 | 14 | 0 |
| Long Write | Grade 11 | Organization | 43479 | 344 | 0 | 25 | 287 | 32 | 0 | 0 | 7 | 83 | 9 | 0 |
| | | Elaboration | 43479 | 344 | 0 | 16 | 290 | 38 | 0 | 0 | 5 | 84 | 11 | 0 |
| | | Conventions | 43479 | 344 | 0 | 17 | 284 | 43 | 0 | 0 | 5 | 83 | 13 | 0 |
| Mathematics | Grade 04 | | 43572 | 90 | 0 | 5 | 85 | 0 | 0 | 0 | 6 | 94 | 0 | 0 |
| Mathematics | Grade 04 | | 43564 | 90 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Mathematics | Grade 04 | | 43173 | 90 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Mathematics | Grade 07 | | 43551 | 90 | 0 | 0 | 87 | 3 | 0 | 0 | 0 | 97 | 3 | 0 |
| Mathematics | Grade 07 | | 43555 | 162 | 0 | 1 | 153 | 8 | 0 | 0 | 1 | 94 | 5 | 0 |
| Mathematics | Grade 07 | | 43557 | 90 | 0 | 1 | 89 | 0 | 0 | 0 | 1 | 99 | 0 | 0 |
| Mathematics | Grade 07 | | 43639 | 99 | 0 | 0 | 99 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Mathematics | Grade 11 | | 43559 | 90 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Mathematics | Grade 11 | | 43552 | 90 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Mathematics | Grade 11 | | 43546 | 90 | 0 | 0 | 90 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Reading | Grade 04 | | 43707 | 110 | 0 | 1 | 108 | 1 | 0 | 0 | 1 | 98 | 1 | 0 |
| Reading | Grade 04 | | 43412 | 110 | 0 | 10 | 94 | 6 | 0 | 0 | 9 | 85 | 5 | 0 |
| Reading | Grade 04 | | 43416 | 78 | 0 | 0 | 78 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Reading | Grade 07 | | 43248 | 101 | 0 | 9 | 84 | 8 | 0 | 0 | 9 | 83 | 8 | 0 |
| Reading | Grade 07 | | 43445 | 110 | 0 | 12 | 92 | 3 | 3 | 0 | 11 | 84 | 3 | 3 |
| Reading | Grade 07 | | 43422 | 110 | 5 | 20 | 80 | 5 | 0 | 5 | 18 | 73 | 5 | 0 |

# Validity Summary Report
## Smarter Balanced Assessment Consortium
## Small Scale Trials

| Subject | Grade | Domain | ID | #R | #NA LO | #LO | #EX | #HI | #NA HI | %NA LO | %LO | %EX | %HI | %NA HI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reading | Grade 11 | | 43297 | 100 | 0 | 6 | 92 | 2 | 0 | 0 | 6 | 92 | 2 | 0 |
| Reading | Grade 11 | | 43435 | 100 | 3 | 10 | 85 | 2 | 0 | 3 | 10 | 85 | 2 | 0 |
| Reading | Grade 11 | | 43397 | 100 | 1 | 16 | 80 | 3 | 0 | 1 | 16 | 80 | 3 | 0 |
| Research | Grade 04 | | 43280 | 100 | 1 | 15 | 81 | 3 | 0 | 1 | 15 | 81 | 3 | 0 |
| Research | Grade 07 | | 43468 | 100 | 1 | 5 | 85 | 9 | 0 | 1 | 5 | 85 | 9 | 0 |
| Research | Grade 11 | | 43491 | 100 | 0 | 19 | 81 | 0 | 0 | 0 | 19 | 81 | 0 | 0 |
| | | | | 10,671 | 24 | 493 | 9,402 | 747 | 5 | 0 | 5 | 88 | 7 | 0 |

## Item Analysis and Data Review Procedures
## Classical Item Analysis

After the automated scoring of the CR responses, the complete dataset was subjected to item analyses.  AIR's analysis program computes individual item and overall test statistics for each selected-response and constructed-response item to check the integrity of the item and to verify the appropriateness of the difficulty level of the item. The score used to compute the biserial correlations and the DIF ability stratification was the raw number-correct score, within form, based on the human scores for the constructed-response (CR) items.

Key statistics that were computed and examined include the following:

**Item Discrimination:** The discrimination index is calculated as the correlation between the item score and the student's total number-correct score (biserial correlations for selected-response items and polyserial correlations for constructed-response items). Selected-response items were flagged for subsequent review if the biserial correlation for the item was less than .25 for the keyed (correct) response and greater than zero for distractors. Constructed-response items were flagged if the polyserial correlation was less than .25.

**Item Difficulty:** Items that are either extremely difficult or extremely easy were flagged for review but not necessarily for removal if the item discrimination index was not also flagged. For selected-response items, the proportion of test-takers in the sample selecting the correct answer (the *p*-value) was computed, as were the proportions of those selecting each incorrect response For constructed-response items, item difficulty was calculated both as the item's mean score (average item score) and as the average proportion correct (analogous to *p*-value). Items were flagged for review if the *p*-value was less than .25 or greater than .90.

Constructed-response items were flagged if there were very few students scoring in a given category or if a very high proportion of students fell in any single score-point category.  The latter may suggest that the other score points are not useful or, if the score point is in the minimum or maximum score-point category, that the item may be too difficult. Constructed-response items were also flagged if the average ability estimate of students in a score-point category was lower than the average ability estimate of students in the next lower score-point category. For example, an item was flagged for review if the average total score of those receiving a score of 3 on the constructed-response item was lower than the average total score for those receiving a 2 on the constructed-response item.

**Differential Item Functioning (DIF):** DIF analyses are designed to determine whether students at similar levels of ability have different probabilities of answering the same item correctly (or of receiving the same scores in the case of constructed-response items) based on group membership. A variety of factors may lead to differential item functioning, but DIF may indicate item bias.

DIF analyses were conducted on all items included in the Small Scale Trial forms to detect potential item bias for subgroups (sample sizes permitting). The performance on each item by focal group members (e.g., protected ethnic group members, females) was compared with the performance of the appropriate reference group (e.g., white students, male students). The purpose of these analyses is to identify items that may have favored students in one group (reference group) over students of similar ability in another group (focal group).

14

The procedures used for detecting DIF are the Mantel-Haenszel (MH, 1959) chi-square for dichotomous items (multiple-choice items) and Mantel's chi-square for polytomous items (constructed-response items). The Mantel-Haenszel statistic (MH D-DIF) is calculated for multiple-choice items (Holland & Thayer, 1988) and standardized mean difference (SMD) for constructed-response items (Dorans & Schmitt, 1991; Zwick, Donoghue, & Grima, 1993) to measure the degree and magnitude of DIF. The total scale score on the test was used as the ability-matching variable. The AIR analysis program computes the MH chi-square value, the log-odds ratio, the standard error of the log-odds ratio, and the MH-delta for the selected-response items, as well as the MH chi-square, the SMD, and the standard error of the SMD for the constructed-response items. Items were classified into three categories (A, B, or C) ranging from no DIF to mild DIF to severe DIF according to the DIF classification convention. Items were also categorized as positive DIF (i.e., A+, B+, or C+), signifying that the item favored the focal group, or negative DIF (i.e., A−, B−, or C−, signifying that the item favored the reference group.

A DIF classification of C means that the item shows significant DIF and should be reviewed for potential content bias, differential validity, or other issues that may reduce item fairness. Items in the C category for any group were flagged for subsequent review by the Fairness Data Review Committee. Table 5 details the DIF classification rules.

Table 5. DIF Classification Rules

| DIF CATEGORY | FLAG CRITERIA |
|---|---|
| **Dichotomous Items** | |
| C | $MH\chi^2$ is significant and $|\hat{\Delta}_{MH}| \geq 1.5$. |
| B | $MH\chi^2$ is significant and $|\hat{\Delta}_{MH}| < 1.5$. |
| A | $MH\chi^2$ is not significant. |
| **Polytomous Items** | |
| C | $MH\chi^2$ is significant and $|SMD|/|SD| \geq .25$. |
| B | $MH\chi^2$ is significant and $|SMD|/|SD| < .25$. |
| A | $MH\chi^2$ is not significant. |

## Item and Test Form Results

Table 6 presents test score (raw) information, by test form, within content area. For reading and writing, the means and standard deviations were similar across grades and forms. The writing forms were of moderate difficulty. The reading forms were a little more challenging. The mathematics forms tended to be the most challenging and tended to get more difficult with increasing grade. No student received the maximum possible points on any test.

Table 6. Raw Score Descriptive Statistics by Test Form

| Content | Grade | Form | N | Mean | SD | Min Obs | Max Obs | Max Possible Points |
|---|---|---|---|---|---|---|---|---|
| Math | 4 | A | 1062 | 8.4 | 3.0 | 1 | 17 | 21 |
| Math | 7 | A | 710 | 6.0 | 3.4 | 0 | 19 | 21 |
| Math | 11 | A | 570 | 4.9 | 2.9 | 0 | 18 | 21 |
| Reading | 4 | A | 1057 | 9.1 | 3.9 | 0 | 19 | 21 |
| Reading | 7 | A | 706 | 8.9 | 3.5 | 1 | 20 | 21 |
| Reading | 11 | A | 580 | 8.3 | 3.4 | 0 | 18 | 22 |
| Writing | 4 | A | 1047 | 9.9 | 4.1 | 0 | 21 | 27 |
| Writing | 4 | B | 1061 | 11.6 | 4.5 | 2 | 23 | 27 |
| Writing | 4 | C | 1042 | 10.1 | 4.4 | 1 | 24 | 27 |
| Writing | 7 | A | 698 | 12.0 | 4.1 | 1 | 26 | 27 |
| Writing | 7 | B | 703 | 12.4 | 4.2 | 2 | 25 | 27 |
| Writing | 7 | C | 695 | 12.4 | 4.5 | 3 | 25 | 27 |
| Writing | 11 | A | 573 | 13.4 | 4.5 | 1 | 23 | 27 |
| Writing | 11 | B | 573 | 12.3 | 4.1 | 2 | 21 | 27 |
| Writing | 11 | C | 576 | 13.0 | 4.7 | 0 | 23 | 27 |

## Item Analysis Results

Item analysis results are reported, by form, in Tables A1–O1 in Appendix C. Two brief writing items (43966, grade 4; 43486, grade 11) were removed prior to analysis after content review. Similar to the test scores, the mean item difficulties for the writing items were of middle difficulty, the reading items were slightly more difficult, and the mathematics forms had more difficult items. The item

discriminations (biserial/polyserial correlations) were quite high, averaging between the middle .40's and the middle .50's.

## DIF Results

To avoid large numbers of false positives, DIF analyses were only run if the sample size in the focal group was at least 100. As a consequence, the primary comparisons were between male/female students and Hispanic/white students. Sample sizes were large enough for some LEP/non-LEP comparisons in grade 4 where the sample sizes were larger. Both positive DIF (favoring the focal group) and negative DIF (favoring the reference group) were observed. Table 7 provides a count of the number of items flagged for DIF and the direction of the DIF. More specific results can be found in Tables A2–O2 in Appendix C. There was nothing unusual about the pattern of DIF in these forms.

Table 7: DIF Flags by Content (Form), Grade, and Comparison

| Grade (Form) | Grade | LEP vs. non-LEP | | Female vs. Male | | Hispanic vs. White | |
|---|---|---|---|---|---|---|---|
| | | C– | C+ | C– | C+ | C– | C+ |
| Mathematics | 4 | - | 2 | - | - | - | 1 |
| Mathematics | 7 | - | - | - | - | - | 1 |
| Mathematics | 11 | - | - | - | - | - | - |
| Reading | 4 | - | - | - | - | - | - |
| Reading | 7 | - | - | 1 | - | 1 | 1 |
| Reading | 11 | - | - | - | - | - | - |
| Writing (Form A) | 4 | - | - | - | - | - | - |
| Writing (Form B) | 4 | - | - | - | - | - | - |
| Writing (Form C) | 4 | - | - | - | - | 1 | - |
| Writing (Form A) | 7 | - | - | - | - | 2 | - |
| Writing (Form B) | 7 | - | - | - | 1 | 1 | - |
| Writing (Form C) | 7 | - | - | - | - | - | - |
| Writing (Form A) | 11 | - | - | 1 | - | 1 | 1 |
| Writing (Form B) | 11 | - | - | 1 | - | - | - |
| Writing (Form C) | 11 | - | - | 1 | - | 2 | 1 |

## Model Building Analyses

To evaluate model building success, the following analyses were performed:

1. a comparison of the descriptive statistics of the item scores under human and automated model scoring;
2. the percentage exact agreement and agreement within 1 point (adjacent match), between the first human score  and the second human score;
3. the percentage exact agreement and agreement within 1 point (adjacent match), between the first human score and the automated model score;
4. the polychoric and Pearson correlations between the validated human score, the first human score and the automated score;
5. Kappa statistics to assess the amount of agreement in scores over chance agreement.


## Development of Automated Scoring Models


## Two Scoring Model Approaches

The two primary classes of automated scoring models examined here are: the empirically developed black-box model, primarily used for scoring long writes/essays, and the theoretically based glass-box models driven by the specificity of the scoring rubric, primarily used for scoring short semantic responses.  Full rubrics for the Small Scale Trial items can be found in Appendix B.

The development of the engines is an iterative process. This is especially true when newly developed models are used. Therefore, following the initial scoring model development, cases where the scoring engine fails to accurately score a response are examined initially on a regular basis and periodically thereafter. This information is used to identify places where the scoring model can be improved. The recommended safety measure is for humans to periodically score a sample of papers for each model and then compare these responses with the automated scores during operational scoring.


*Automated Essay Scoring Model (Black-Box Model)*

The development of an automated essay scoring model is a data-driven approach. While black-box essay scoring engines may correlate reasonably well with human scores, they only incorporate shallow semantics.  They do not evaluate the logic or quality of argumentation.  These elements are within the scope of the Consortium scoring rubrics. For example, the "Evidence and Elaboration" domain for the argumentative rubric contains the specifications identified in Exhibit 1.

Exhibit 1: Evidence and Elaboration
4 point score description, Argumentation

The response provides thorough and convincing support/evidence for the controlling idea or main idea that includes the effective use of sources, facts, and details. The response clearly and effectively expresses ideas, using precise language:

- comprehensive evidence from sources is integrated

- references are relevant and specific

- effective use of a variety of elaborative techniques is demonstrated

Use of domain-specific vocabulary is clearly appropriate for the audience and purpose.

Essay scoring engines can pick up on relevance of the vocabulary to the topic in question, but have no access to the logic, other than through the incidental correlations with vocabulary or other syntactic features of the text. Proposition scoring engines cannot currently score such responses because the number and complexity of possible reasonable arguments cannot realistically be enumerated. Hence, the valid scoring of such rubric elements extends beyond the current state of the art.

That said, the naturally occurring correlations between the features these engines measure and writing traits that we want to measure often results in accurate scoring in many cases. It is the less common responses or the responses with good argumentation but (for example) poor spelling or word choice where the correlations would not prove accurate predictors.

## Human Scoring for Model Training

Black-box models are developed in two phases, a training phase and a validation phase. The training for black-box scoring models uses human ratings as the primary source of information for developing the automated scoring model. As such, the quality of the automated model is related to the level of agreement between the human raters. This will be reflected in the consistency of the scores assigned. To provide the best information for the automated model development process, it is

important that the human ratings be as accurate as possible. The best way to accomplish this is to have each student response scored twice by two trained human scorers. If the scores for two trained scorers do not match exactly, the scores should be sent to an expert scorer for resolution. The score used in the model building should include the information from this resolution process.

During the training phase, the scoring engine is subjected to exemplars that define the bounds within which to recognize patterns. This is a data-driven approach, so human scoring needs to have been completed prior to the development of the model. This methodology is applied to responses that are varied and require complex modeling, including the extraction of syntactic and semantic features, followed by feature-space mapping and dimension reduction approaches, followed in turn by regression or other statistical prediction of the validated human scores based on the dimension-reduced feature-space.

### Propositional Scoring (Glass Box) Model

The proposition scoring engine is a glass-box model for which test developers build explicit rubrics. This model uses a set of (potentially interrelated) propositions.  This approach differs from natural language understanding in that it seeks to recognize relationships specified by the rubric author, rather than to infer relationships from natural language.

Broadly speaking, propositions are built from concepts and relationships. A concept is a collection of words that have similar meaning.  Similarity may be defined as synonymy, ontological relationships, or other relationships that may be selected by the rubric author. Concepts may be modified by specifying their scope.  For example, *a dog* refers to any instance of a dog, while *the dog* refers to a specific instance of a dog, and *dogs* will refer to the abstract entity rather than any single instance.

Relationships are represented using triplets, containing an agent, an object, and a relationship. Each element in the triplet may be modified or expanded by attributes.  The triplets represent syntax-independent descriptions of concepts. In most cases, relations correspond to verbs.  For example,

*A dog chased a cat*, or *the dog chased a cat*, or *the cat was chased by the dog* are all represented by the relational triplet in Figure 1.
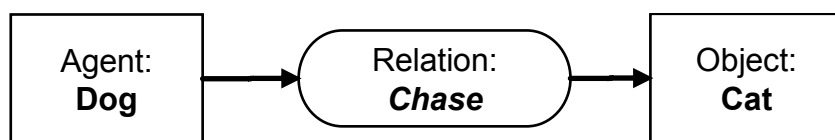


Figure 1: Relational Concept representing "the dog chased the cat"

The sample sentences might be distinguished by adding qualifiers.  For instance, specifying that the dog was a named instance might preclude a match in a sentence referring to *a* dog.

The determination of equivalence concepts is useful for defining the synonymy between concepts and their relational equivalents.  Continuing the example above, *the dog treed a cat* is just another way of saying that the dog chased the cat up the tree. Recognizing the equivalence, however,

requires the semantic knowledge that to "tree" means to "chase up a tree." Equivalence concepts provide a mechanism for encoding this semantic information.

When an examinee response is captured, it is first parsed into a syntactic parse tree (like a sentence diagram), using a parsing algorithm. The matching algorithm then searches the parse tree for evidence of the propositions defined in the rubric. A final scoring stage assigns scores based on Boolean collections of propositions.

Test developers build the concepts to be scored in student responses through specifications that fit within the specific scoring engine template based on the scoring rubric for the item type. The validation of glass-box models entails *rubric validation* to verify that the logic and salient features of the rubric are complete. When selecting papers for rubric validation, it is useful to disproportionately select high scores on the target items received by otherwise low-scoring students and low scores on the target items received by otherwise high-scoring students as a successful strategy to refine the machine-scored rubric.

Items that are intended to be scored with a glass-box approach will be automatically scored using the machine-scored rubric created during item development. These items will have preexisting scores that will be validated during the range finding/rubric validation process. The items will have a range of correct responses, and the range finding/rubric validation committee will validate or adjust the computer-generated scores or broaden the scoring rubric to encompass additional valid responses.

## Scoring Model Results

### Essay (black box) model results

The long-write essay items from the writing test forms were scored with rubrics for three domains (organization, elaboration, and conventions). Separate scoring models were developed for each domain. Descriptive information about sample size, average essay response length, first and second human scores, and automated scores can be found in the first 4 sections of Table 8. The last two sections of the table compare human/human scores and human/automated scores, respectively. It was planned to obtain 1000 responses per item using 500 responses to train the scoring engine and 500 responses to validate the model. In most cases this was achieved; however, the observed sample sizes were smaller in grade 11. When the sample sizes were small between 400 and 450 responses were used to train the engine and the rest were used to validate the model. This sometimes resulted in small numbers of cases for validation. However, it is important to build a stable model or the validation results will suffer as a consequence.

Essay response length tended to increase by grade, with grade 4 students producing noticeably shorter responses. For the organization and elaboration scores, the means and standard deviations were very similar between the human and automated scores. Agreement and correlational measures tended to be slightly lower for the human/automated values than for the human/human indices for these measures. Two cross-tabulated agreement tables were constructed for each item: one compared the first human score with the second human score; the other compared the first human score with the score produced by the automated scoring engine. The human/human table should be used as a baseline against which the human/automated results are compared. The tables containing these results can be found in Appendix D. The agreement between the

human/automated scoring and the human/human scoring was uniformly high for the organization and elaboration rubrics, though the human/automated indices were always lower.

For the conventions rubric, the results are less consistent, particularly for the fourth grade items. The means were further apart and the standard deviations for the automated model were uniformly smaller.  Agreement and correlational measures were substantially lower for the human/automated comparisons than they were for the human/human comparisons.  The agreement for conventions between human/human comparisons was lower than expected and an issue for which research is ongoing. The automated model score distributions, shown in Appendix D, tended to be more peaked than the human distribution.  There were fewer extreme scores assigned by the automated model (regression effect).

## Propositional (glass box) model results

For constructed-response items, descriptive information about sample size, average response length, first and second human scores, and automated scores can be found in the first four sections of Table 9. The last two sections compare human/human scores and human/automated scores, respectively. The sample size issues outlined above hold for these items as well. With one exception, 501 responses were used to hone the propositional model rubrics. For item 43564, 471 responses were used for this purpose. The response length for these items tended to be between 20 and 40 words.  Writing items in grades 7 and 11 tended to be a little longer, mathematics items in grade 4 tended to be between 10 and 15 words.

The validity standard used to qualify raters was an exact agreement of 90% for a 1-point item and 80% for a 2-point item. Overall, the performance of the scoring engine met this standard for 50% of the items across all grades and subjects.

*Mathematics*

About 67% of the cases (six out of nine items) met the validity standard used to qualify raters.  The scoring engine performed well enough for six items, including the two cases in which the engine performed exactly as well as humans. The cross-tabulated agreement tables, shown in Appendix E, compare the first human score with the second human score and the first human score with the automated score.  Overall, agreement and correlational measures for the human/automated relationship tended to range from the same as or slightly lower than the human/human indices for these measures. The cross-tabulated agreement tables, shown in Appendix E, compare the first human score with the second human score and the first human score with the automated score. As indicated in the tables, more low or 0 scores were produced using the propositional scoring model.

There was an issue with some mathematics items, unrelated to the scoring model, in that all or almost all students received a score of 0 from human scorers. This was true for item 43564 in grade 4, and items 43559 and 43552 in grade 11.

*Reading*

The reading responses were about the same length across grades. For only 33% of the items, the engine performed well enough to meet the validity standard. For the items in which the engine failed to meet the standard, the exact agreements between human scorers tended to be about 10% lower than in 'met-standard' cases. This may suggest that the rubrics were not as clearly specified as needed for these items. Overall, agreement and correlational measures tended to be lower for the human/automated score comparison than for the human/human score comparison. The cross-tabulated agreement tables comparing the first human score with the second human score and the first human score with the score produced by the automated scoring engine indicate that the automated scoring engine produced lower scores with substantial numbers of zero scores. The agreement tables showing this can be found in Appendix E.

*Writing*

The writing items where the propositional model was applied were limited in number, with only one item per grade. The engine performed well for two of the three items. The engine met the validity standard for the grades 4 and 7 items. The grade 7 item showed not only good exact agreement but comparable means between the human and automated scores. The standard deviations for the automated model were smaller than or similar to the human score for the grades 4, 7, and 11 items. The agreement and correlational measures were lower for the human/automated score comparison than for the human/human score comparison, particularly for the grade 11 item. The cross-tabulated agreement tables comparing the first human score with the second human score and the first human score with the score produced by the automated scoring engine support these results. The agreement tables showing this can be found in Appendix E.

Table 8. Statistics Summary for the Multiple Dimension Essay Writing Items

| Item ID | Dimension | Grade | Model Validation N | Model Training N | Mean – Word Count | First Human Scorer | | | | Second Human Scorer | | | | Automated Score | | | | First vs Second Human Scorers | | | | | First Human Scorer vs Automated Score | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum | Exact Agreement | Agreement Within 1pt | Pearson Correlation | Polychoric/Tetrachoric Correlation | Weight Kappa | Exact Agreement | Agreement Within 1pt | Pearson Correlation | Polychoric/Tetrachoric Correlation | Weighted Kappa |
| 43504 | Organization | 4 | 439 | 425 | 41 | 1.47 | 0.61 | 1 | 4 | 1.46 | 0.61 | 1 | 4 | 1.42 | 0.55 | 1 | 3 | 86 | 100 | 0.81 | 0.92 | 0.81 | 78 | 100 | 0.68 | 0.85 | 0.67 |
| | Elaboration | 4 | 439 | 425 | 41 | 1.47 | 0.6 | 1 | 4 | 1.45 | 0.58 | 1 | 4 | 1.46 | 0.56 | 1 | 3 | 86 | 100 | 0.8 | 0.93 | 0.8 | 80 | 100 | 0.71 | 0.87 | 0.7 |
| | Convention | 4 | 439 | 425 | 41 | 1 | 0.79 | 0 | 2 | 1.03 | 0.79 | 0 | 2 | 1.04 | 0.63 | 0 | 2 | 76 | 100 | 0.79 | 0.9 | 0.79 | 46 | 98 | 0.41 | 0.51 | 0.4 |
| 43334 | Organization | 4 | 485 | 451 | 52 | 1.64 | 0.66 | 1 | 4 | 1.64 | 0.66 | 1 | 4 | 1.61 | 0.67 | 1 | 4 | 83 | 100 | 0.81 | 0.94 | 0.81 | 83 | 100 | 0.79 | 0.9 | 0.79 |
| | Elaboration | 4 | 485 | 451 | 52 | 1.62 | 0.64 | 1 | 3 | 1.63 | 0.65 | 1 | 3 | 1.6 | 0.66 | 1 | 4 | 82 | 100 | 0.79 | 0.92 | 0.79 | 83 | 99 | 0.78 | 0.89 | 0.78 |
| | Convention | 4 | 485 | 451 | 52 | 0.87 | 0.77 | 0 | 2 | 0.91 | 0.78 | 0 | 2 | 0.92 | 0.65 | 0 | 2 | 81 | 100 | 0.84 | 0.94 | 0.84 | 51 | 98 | 0.46 | 0.57 | 0.45 |
| 43284 | Organization | 4 | 396 | 394 | 55 | 1.6 | 0.73 | 1 | 4 | 1.6 | 0.72 | 1 | 4 | 1.61 | 0.72 | 1 | 4 | 80 | 100 | 0.81 | 0.92 | 0.81 | 77 | 99 | 0.76 | 0.88 | 0.76 |
| | Elaboration | 4 | 396 | 394 | 55 | 1.6 | 0.73 | 1 | 4 | 1.61 | 0.71 | 1 | 4 | 1.6 | 0.7 | 1 | 4 | 80 | 100 | 0.81 | 0.93 | 0.81 | 77 | 100 | 0.77 | 0.89 | 0.77 |
| | Convention | 4 | 396 | 394 | 55 | 0.84 | 0.76 | 0 | 2 | 0.85 | 0.77 | 0 | 2 | 0.74 | 0.7 | 0 | 2 | 81 | 99 | 0.82 | 0.92 | 0.82 | 47 | 93 | 0.32 | 0.39 | 0.32 |
| 43438 | Organization | 7 | 164 | 442 | 102 | 1.59 | 0.62 | 1 | 3 | 1.54 | 0.6 | 1 | 3 | 1.59 | 0.63 | 1 | 3 | 82 | 99 | 0.71 | 0.84 | 0.71 | 74 | 100 | 0.66 | 0.82 | 0.66 |
| | Elaboration | 7 | 164 | 442 | 102 | 1.57 | 0.62 | 1 | 3 | 1.52 | 0.6 | 1 | 3 | 1.55 | 0.62 | 1 | 3 | 82 | 99 | 0.71 | 0.83 | 0.7 | 77 | 100 | 0.7 | 0.86 | 0.7 |
| | Convention | 7 | 164 | 442 | 102 | 1.04 | 0.81 | 0 | 2 | 1.09 | 0.76 | 0 | 2 | 1.23 | 0.69 | 0 | 2 | 69 | 99 | 0.73 | 0.85 | 0.73 | 52 | 96 | 0.49 | 0.6 | 0.47 |
| 43703 | Organization | 7 | 175 | 441 | 107 | 1.39 | 0.53 | 1 | 3 | 1.37 | 0.54 | 1 | 3 | 1.39 | 0.5 | 1 | 3 | 83 | 100 | 0.7 | 0.88 | 0.7 | 82 | 100 | 0.67 | 0.86 | 0.67 |
| | Elaboration | 7 | 175 | 443 | 107 | 1.37 | 0.51 | 1 | 3 | 1.37 | 0.53 | 1 | 3 | 1.4 | 0.5 | 1 | 3 | 85 | 100 | 0.71 | 0.89 | 0.71 | 83 | 100 | 0.68 | 0.87 | 0.67 |
| | Convention | 7 | 176 | 441 | 108 | 0.93 | 0.8 | 0 | 2 | 0.9 | 0.81 | 0 | 2 | 0.88 | 0.77 | 0 | 2 | 78 | 99 | 0.81 | 0.91 | 0.81 | 53 | 96 | 0.53 | 0.63 | 0.53 |
| 43469 | Organization | 7 | 153 | 432 | 147 | 1.62 | 0.73 | 1 | 4 | 1.61 | 0.72 | 1 | 4 | 1.59 | 0.7 | 1 | 4 | 82 | 100 | 0.83 | 0.93 | 0.83 | 82 | 100 | 0.83 | 0.94 | 0.83 |
| | Elaboration | 7 | 153 | 448 | 147 | 1.59 | 0.73 | 1 | 4 | 1.58 | 0.71 | 1 | 4 | 1.55 | 0.68 | 1 | 4 | 80 | 100 | 0.81 | 0.92 | 0.81 | 82 | 100 | 0.83 | 0.94 | 0.82 |
| | Convention | 7 | 156 | 432 | 144 | 0.85 | 0.75 | 0 | 2 | 0.79 | 0.73 | 0 | 2 | 0.72 | 0.73 | 0 | 2 | 78 | 100 | 0.8 | 0.92 | 0.8 | 62 | 97 | 0.59 | 0.71 | 0.59 |
| 43632 | Organization | 11 | 64 | 444 | 173 | 1.89 | 0.67 | 1 | 3 | 1.97 | 0.69 | 1 | 3 | 1.88 | 0.63 | 1 | 3 | 86 | 100 | 0.85 | 0.97 | 0.85 | 86 | 100 | 0.83 | 0.96 | 0.83 |
| | Elaboration | 11 | 64 | 444 | 173 | 1.84 | 0.65 | 1 | 3 | 1.91 | 0.66 | 1 | 3 | 1.8 | 0.67 | 1 | 3 | 84 | 100 | 0.82 | 0.95 | 0.82 | 83 | 100 | 0.8 | 0.94 | 0.8 |
| | Convention | 11 | 64 | 444 | 173 | 1.39 | 0.68 | 0 | 2 | 1.39 | 0.68 | 0 | 2 | 1.38 | 0.7 | 0 | 2 | 81 | 100 | 0.79 | 0.92 | 0.79 | 59 | 95 | 0.42 | 0.53 | 0.42 |
| 43635 | Organization | 11 | 65 | 444 | 224 | 2.09 | 0.74 | 1 | 3 | 2.05 | 0.76 | 1 | 3 | 2.03 | 0.75 | 1 | 3 | 83 | 100 | 0.85 | 0.95 | 0.85 | 82 | 100 | 0.84 | 0.94 | 0.83 |
| | Elaboration | 11 | 65 | 444 | 224 | 2.02 | 0.7 | 1 | 3 | 2.02 | 0.74 | 1 | 3 | 2.05 | 0.74 | 1 | 3 | 88 | 100 | 0.88 | 0.98 | 0.88 | 85 | 100 | 0.85 | 0.96 | 0.85 |

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Convention | 11 | 65 | 444 | 224 | 1.2 | 0.71 | 0 | 2 | 1.26 | 0.71 | 0 | 2 | 1.46 | 0.59 | 0 | 2 | 72 | 100 | 0.73 | 0.86 | 0.72 | 57 | 98 | 0.52 | 0.66 | 0.47 |
| | Organization | 11 | 69 | 455 | 164 | 1.75 | 0.63 | 1 | 3 | 1.84 | 0.68 | 1 | 3 | 1.77 | 0.67 | 1 | 3 | 88 | 100 | 0.87 | 0.99 | 0.86 | 81 | 100 | 0.77 | 0.92 | 0.77 |
| 43479 | Elaboration | 11 | 69 | 455 | 164 | 1.71 | 0.67 | 1 | 3 | 1.7 | 0.6 | 1 | 3 | 1.62 | 0.64 | 1 | 3 | 84 | 100 | 0.8 | 0.95 | 0.8 | 77 | 100 | 0.74 | 0.88 | 0.73 |
| | Convention | 11 | 69 | 455 | 164 | 1.3 | 0.71 | 0 | 2 | 1.2 | 0.7 | 0 | 2 | 1.25 | 0.69 | 0 | 2 | 75 | 100 | 0.76 | 0.9 | 0.75 | 55 | 99 | 0.5 | 0.61 | 0.5 |

Table 9. Statistics Summary of Propositional Model

| Item ID | Form | N | Mean - Word Count | First Human Scorer | | | | Second Human Scorer | | | | Automated Score | | | | First vs. Second Human Scorers | | | | | | First Human Scorer vs. Automated Score | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum | Mean | Standard Deviation | Minimum | Maximum | Exact Agreement | Meet Validity Standard? (1pt-90%; 2pt- 80%) | Agreement Within 1pt | Pearson Correlation | Polychoric/Tetrachoric Correlation | Weight Kappa | Exact Agreement | Meet Validity Standard? (1pt-90%; 2pt- 80%) | Agreement Within 1pt | Pearson Correlation | Polychoric/Tetrachoric Correlation | Weighted Kappa |
| 43572 | Math G4 | 532 | 14 | 0.41 | 0.49 | 0 | 1 | 0.41 | 0.49 | 0 | 1 | 0.24 | 0.43 | 0 | 1 | 95 | Yes | 100 | 0.90 | 0.99 | 0.90 | 74 | | 100 | 0.45 | 0.70 | 0.42 |
| 43564 | Math G4 | 156 | 10 | 0.01 | 0.08 | 0 | 1 | 0.01 | 0.08 | 0 | 1 | 0.01 | 0.08 | 0 | 1 | 100 | Yes | 100 | 1.00 | 1.00 | 1.00 | 100 | Yes | 100 | 1.00 | 1.00 | 1.00 |
| 43173 | Math G4 | 524 | 12 | 0.15 | 0.36 | 0 | 2 | 0.15 | 0.37 | 0 | 2 | 0.08 | 0.28 | 0 | 2 | 98 | Yes | 100 | 0.91 | 0.99 | 0.91 | 93 | Yes | 100 | 0.67 | 0.91 | 0.64 |
| 43551 | Math G7 | 208 | 25 | 0.24 | 0.43 | 0 | 1 | 0.21 | 0.41 | 0 | 1 | 0.15 | 0.36 | 0 | 1 | 91 | Yes | 100 | 0.75 | 0.94 | 0.75 | 81 | | 100 | 0.42 | 0.68 | 0.41 |
| 43555 | Math G7 | 192 | 25 | 0.42 | 0.52 | 0 | 2 | 0.41 | 0.51 | 0 | 2 | 0.34 | 0.48 | 0 | 1 | 92 | Yes | 100 | 0.84 | 0.97 | 0.84 | 70 | | 99 | 0.37 | 0.53 | 0.36 |
| 43557 | Math G7 | 189 | 20 | 0.32 | 0.47 | 0 | 1 | 0.31 | 0.46 | 0 | 1 | 0.27 | 0.45 | 0 | 1 | 96 | Yes | 100 | 0.90 | 0.99 | 0.90 | 85 | Yes | 100 | 0.65 | 0.87 | 0.65 |
| 43639 | Math G7 | 186 | 19 | 0.37 | 0.48 | 0 | 1 | 0.36 | 0.48 | 0 | 1 | 0.28 | 0.45 | 0 | 1 | 95 | Yes | 100 | 0.90 | 0.99 | 0.90 | 89 | Yes | 100 | 0.76 | 0.95 | 0.74 |
| 43559 | Math G11 | 54 | 24 | 0.04 | 0.19 | 0 | 1 | 0.02 | 0.14 | 0 | 1 | 0.00 | 0.00 | 0 | 0 | 98 | Yes | 100 | 0.70 | 1.00 | 0.66 | 96 | Yes | 100 | N/A | N/A | 0.00 |
| 43552 | Math G11 | 24 | 21 | 0.00 | 0.00 | 0 | 0 | 0.00 | 0.00 | 0 | 0 | 0.00 | 0.00 | 0 | 0 | 100 | Yes | 100 | N/A | N/A | N/A | 100 | Yes | 100 | N/A | N/A | N/A |
| 43707 | Read G4 | 540 | 26 | 0.65 | 0.80 | 0 | 2 | 0.67 | 0.82 | 0 | 2 | 0.28 | 0.59 | 0 | 2 | 91 | Yes | 98 | 0.88 | 0.60 | 0.88 | 71 | | 89 | 0.53 | 0.78 | 0.44 |
| 43412 | Reading G4 | 524 | 19 | 0.41 | 0.58 | 0 | 2 | 0.41 | 0.57 | 0 | 2 | 0.28 | 0.52 | 0 | 2 | 90 | Yes | 100 | 0.86 | 0.97 | 0.86 | 81 | Yes | 99 | 0.68 | 0.87 | 0.66 |
| 43416 | Reading G4 | 537 | 15 | 0.74 | 0.75 | 0 | 2 | 0.73 | 0.73 | 0 | 2 | 0.66 | 0.66 | 0 | 2 | 93 | Yes | 100 | 0.92 | 0.97 | 0.92 | 84 | Yes | 99 | 0.84 | 0.93 | 0.83 |
| 43248 | Reading G7 | 208 | 33 | 0.50 | 0.56 | 0 | 2 | 0.53 | 0.60 | 0 | 2 | 0.09 | 0.31 | 0 | 2 | 77 | | 100 | 0.64 | 0.80 | 0.64 | 59 | | 98 | 0.27 | 0.51 | 0.16 |
| 43445 | Reading G7 | 186 | 36 | 0.46 | 0.71 | 0 | 2 | 0.41 | 0.66 | 0 | 2 | 0.16 | 0.41 | 0 | 2 | 80 | Yes | 97 | 0.70 | 0.84 | 0.69 | 69 | | 94 | 0.43 | 0.64 | 0.33 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43422 | Reading G7 | 188 | 41 | 0.48 | 0.73 | 0 | 2 | 0.46 | 0.71 | 0 | 2 | 0.11 | 0.37 | 0 | 2 | 84 | Yes | 94 | 0.68 | 0.84 | 0.67 | 70 | | 90 | 0.40 | 0.69 | 0.27 |
| 43297 | Reading G11 | 60 | 37 | 0.58 | 0.67 | 0 | 2 | 0.58 | 0.72 | 0 | 2 | 0.28 | 0.52 | 0 | 2 | 83 | Yes | 100 | 0.83 | 0.95 | 0.83 | 72 | | 98 | 0.68 | 0.89 | 0.58 |
| 43435 | Reading G11 | 64 | 34 | 0.41 | 0.61 | 0 | 2 | 0.25 | 0.47 | 0 | 2 | 0.20 | 0.51 | 0 | 2 | 81 | Yes | 97 | 0.58 | 0.78 | 0.54 | 70 | | 97 | 0.45 | 0.66 | 0.41 |
| 43397 | Reading G11 | 39 | 33 | 0.33 | 0.70 | 0 | 2 | 0.28 | 0.65 | 0 | 2 | 0.23 | 0.43 | 0 | 1 | 97 | Yes | 97 | 0.89 | 0.98 | 0.88 | 79 | Yes | 100 | 0.79 | 0.96 | 0.69 |
| 43280 | Writing G4 | 485 | 32 | 0.33 | 0.61 | 0 | 2 | 0.33 | 0.58 | 0 | 2 | 0.19 | 0.45 | 0 | 2 | 92 | Yes | 100 | 0.87 | 0.97 | 0.87 | 81 | Yes | 98 | 0.63 | 0.82 | 0.58 |
| 43468 | Writing G7 | 165 | 49 | 0.79 | 0.83 | 0 | 2 | 0.81 | 0.85 | 0 | 2 | 0.78 | 0.84 | 0 | 2 | 96 | Yes | 100 | 0.97 | 1.00 | 0.97 | 82 | Yes | 98 | 0.83 | 0.92 | 0.83 |
| 43491 | Writing G11 | 55 | 68 | 1.16 | 0.74 | 0 | 2 | 1.16 | 0.76 | 0 | 2 | 0.65 | 0.75 | 0 | 2 | 89 | Yes | 100 | 0.90 | 0.98 | 0.90 | 49 | | 89 | 0.47 | 0.61 | 0.38 |

## Conclusions

*Essay (Black Box) Model*

The primary focus of the Small Scale Trials was on the development and application of two automated scoring models, the essay (black box) scoring model and the propositional (glass box) scoring model. The essay (black box) scoring model was applied separately to scoring rubrics for three domains (organization, elaboration, and conventions). Separate scoring models were developed for each.

The organization and elaboration models performed as well as any automated scoring models presently available. The means and standard deviations were very similar between the human and automated scores. Agreement and correlational measures tended to be slightly lower for the human/automated values than for the human/human indices for these measures, but this is not unusual for automated essay scoring models.

For the conventions rubric, the results were less consistent, particularly for the fourth grade items. There was less agreement in the means, and the standard deviations for the automated model were uniformly smaller. Agreement and correlational measures were substantially lower for the human/automated comparisons than for the human/human comparisons. The automated model score distributions tended to be more peaked than the human distributions. There were fewer extreme scores assigned by the automated model (regression effect). The exact reason for this is unclear at the moment. The conventions rubric was shorter (3 points) versus the longer score scale for the organization and elaboration rubrics (5 points). The distributions clearly show that the automated model was conservative, tending to produce the middle score. These factors could have contributed to the poor performance with the conventions rubric.

The inter-rater reliability between the human to human scores for the conventions domain was lower than expected. This issue is currently being researched.

*Propositional (Glass Box) Model*

The propositional scoring model performed well enough to meet the validity standard for 50% of the constructed-response items across all grades and subjects.

*Mathematics*

The scoring engine met the validity standard for 67% of the items, though the means for the automated model tended to be lower than those for the human produced scores. The standard deviations were very similar to those obtained by human scoring, only slightly smaller. Agreement and correlational measures for the human/automated relationship tended to range from the same as or slightly lower than to substantially lower than the human/human values for these indices.

*Reading*

The reading responses were about the same length across grades. In one-third of all the items, the engine's performance met the validity standard. This low proportion may be partly due to the items themselves because even the exact agreements between human scorers in most 'not-met-standard'

cases were relatively much lower. Agreement and correlational measures tended to be lower for the human/automated score comparison than for the human/human score comparison.

*Writing*

The writing items were limited in number, with only one item per grade. The engine's performance met the validity standard for two of the three items.  The grade 7 item showed not only good exact agreement but comparable means between the human and automated scores. The standard deviations for the automated model were smaller than or similar to those provided by the human raters. The agreement and correlational measures were lower for the human/automated score comparison than for the human/human score comparison, particularly for the grade 4 and 11 items.

Additional work is needed to better understand and improve the propositional model.  Models that are content-based and try to tap into the semantics of an item are a larger challenge than the black box model that relies on structural components in an essay or long writing item. The responses tend to be shorter, often only two or three sentences, which limits the information available for model development. The Small Scale Trial study is only the first look at the application of these types of models.  The information learned here will be applied to the pilot and field test items in an effort to improve the scoring of these types of items.

## Automated Scoring Support

The automated scoring software has been developed to work as well as any engine currently available from any source. This level of assurance does not mean that all performance tasks can be scored using automated models. Efforts will continue to be made to expand this capability in order to reach the Consortium goal of providing authentic assessments that are scored in real time.

The pilot test will provide information about which types of items can currently be scored automatically and which types cannot.  It should also yield information about how consistent a specific scoring model is in scoring. Decisions will have to be made as to whether an automated scoring model is sufficient to provide the final score or if some type of human scoring support is desired.

There is often interest in providing a "safety net" when automated scoring models are used for operational assessments. One answer is to double-score all student responses, first with an automated model, often providing a score in the field, and then, soon after, with a single human score provided through a distributed scoring network as a check. If differences are found between the two scores, the human score can become the official score or the response can be routed to a master grader to provide the official score. This effectively cuts the cost and time of a complete human double-scoring and allows preliminary scores to be reported in the field. It also provides added support to a single human scoring model because two scores are generated.

A further refinement to this model is to limit the number of cases referred to human scoring by using prediction models to identify cases where the score provided by the automated scoring engine is not in keeping with the student's performance on the rest of the test. Cases where the score provided by the automated scoring model falls outside of some tolerance band based on a predicted item score would be referred to a human for score verification. However, cases where the automated model score is in keeping with the predicted score based on the student's performance on the remainder of the test would not be referred for human scoring. This process can be further refined to refer cases

for human scoring only if a possible change in score would change the classification status of the student.

Another means to support the scoring system is to route cases to human scorers only if there are significant "person fit" issues. Person-fit flagging occurs when, for example, a person answers difficult items correctly but answers easy items incorrectly.

### Field Test/Operational Scoring Plan

Given the present state of the art of automated scoring, it is anticipated that the operational scoring model should use automated scoring of some sort for most items, but selectively target those responses that were most tenuously scored to human scorers. We expect that many constructed response item types will be scoreable with sufficient confidence that targeted human backreads will prove unnecessary. In particular, equation response and graphic response items are likely to score with sufficient confidence.

Textual responses are the most challenging to score, using either black-box or glass-box engines. This challenge arises in part from the diversity of language use, and in part from the less explicit, objective criteria for correctness. Essay length responses can often be scored as accurately by scoring engines as by human readers, at least for some types of rubrics; however, as mentioned above, the approach fails in important cases. Shorter responses that have explicitly enumerated correct answers can often be scored accurately, though generally not quite as accurately as by well-trained human readers. Automated scoring approaches tend to perform less well on short texts scored on less explicit rubrics.

It may be possible for the Consortium to develop improved engines for scoring brief writes. The approach might include integrating black-box and glass-box approaches, while at the same time making the rubrics more item specific and explicit.

With these text-processing engines in hand, the Consortium can begin to craft a specific approach to scoring. To support any kind of performance task with a black box or glass box model, it is recommended that a process of automatically identifying "suspicious responses" be put in place. By "suspicious," we mean responses that are more likely to have been mis-scored by the automated engine. These cases, as well as other random cases, periodically sampled, should be routed to human scores for back-reading.

# References

Attali, Y., Powers, D., Freedman, M., Harrison, M., & Obetz, S. (2008). *Automated scoring of short-answer open-ended GRE subject test items.* ETS GRE Board Research Report No. 04-02, ETS RR-08-20. Princeton, NJ: ETS.

Bennett, R. E., Steffen, M., Singley, M. K., Morley, M., & Jacquemin, D. (1997). Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests. *Journal of Educational Measurement, 34*(2), 162–176.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed-response and differential item functioning: A pragmatic approach* (ETS Research Report No. 91-47). Princeton, NJ: Educational Testing Service. [Also appears in *Construction vs. Choice in Cognitive Measurement* (pp. 135-166), R. E. Bennett & W. C. Ward (Eds.), 1993, Hillsdale, NJ: Erlbaum.]

Holland, P. W. (1985). On the study of differential item difficulty without IRT. *Proceedings of the Military Testing Association.*

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity.* Hillsdale, NJ: Erlbaum.

Klein, S. P. (2008). Characteristics of hand and machine-assigned scores to college students' answers to open-ended tasks. In D. Nolan & T. Speed (Eds.), *Probability and statistics: Essays in honor of David A. Freedman*, Vol. 2 (pp. 76–89). Beachwood, OH: Institute of Mathematical Statistics.

Williamson, D. M., Bejar, I. I., & Hone, A. S. (1999). "Mental model" comparison of automated and human scoring. *Journal of Educational Measurement, 36*(2), 158–184.

Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., & Bhola, D. S. (2002). A review of strategies for validating computer-automated scoring. *Applied Measurement in Education, 15*(4), 391–412.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of Differential Item Functioning for Performance Tasks. *Journal of Educational Measurement, 30*(3), 233-251.

# AMERICAN INSTITUTES FOR RESEARCH®

| | |
|---|---|
| **TO:** | Smarter Balanced Assessment Consortium |
| **FROM:** | AIR Technical Team |
| **RE:** | Sampling Plan for SBAC 2012 Small-Scale Trials |
| **DATE:** | August 29, 2012 |

This memo describes the sampling plan for the SBAC small scale trials. If you have questions please contact Gary W. Phillips by email at gwphillips@air.org or by phone at (202) 403-6916.

Due to tight schedules and budget considerations the 2012 SBAC small-scale trials will be scaled back. It has been decided by SBAC that most of the planned research associated with the small-scale trials should move to the pilot test component. However, one piece that cannot be moved is the piece that informs the automated scoring strategy.

Much of the research that had been planned for the small-scale trials was designed to generate evidence of the construct validity of the test and the impact of accommodations on the construct validity. The likely outcome of this research, if validity problems are identified, would be to change the presentation of items, eliminate certain classes of items from the pool, or alter the way that accommodations are implemented.

However, information from the pilot test will begin to become available in May, 2013 (about halfway through the expected item development for the field test), and information about the automated scoring will not become available until August. Given that content and fairness committee meetings are usually held in the summer and that two-thirds of the test development period will have passed by then, it is probably too late for an overhaul of the scoring rubrics and machine-scored constructed-response items. If these data can flow from the trials, they begin to become available in May, about halfway through the field-test development period. While this schedule is still tight and risks some rework, getting automated scoring information from the small-scale trials offers the best chance for success.

The small-scale trial will be limited to grades 4, 7 and 11. Within each grade, a separate 15-item test consisting of selected-response and well-known technology enhanced (TE) item types will be constructed for reading and math. These selected-response item sets will provide a way to evaluate the constructed-response score correlations with the overall test. These data will also be used in strategies to detect anomalous responses.

## Forms Design

AIR will create 15 selected-response (SR) and three constructed-response (CR) items (intended for machine-scoring) in each content area (reading, writing and math) at each grade level (grades 4, 7 and 11). Each participating student would take one of the following forms at the appropriate grade level:

- Form 1: Reading - 15 SR and 3 CR

- Form 2: Writing
    - o   15 SR (9 reading SR and 6 writing SR)
    - o   Brief writing CR A, writing prompt X

- Form 3: Writing
    - o   15 SR (9 reading SR and 6 writing SR)
    - o   Brief writing CR B, writing prompt Y

- Form 4: Writing
    - o   15 SR (9 reading SR and 6 writing SR)
    - o   Research CR C, writing prompt Z

- Form 5: Mathematics - 15 SR and 3 CR


## Sampling Design

The current plan is to administer the small-scale trial in October 2012. AIR will draw a two-stage stratified random sample in grades 4, 7 and 11. Three independent samples will be drawn, one for each grade. The first stage will be a sample of schools that is representative of 25 SBAC states. Within each state a random sample of schools will be obtained that is proportional to the number of schools within the state. We will sample a 20% overage in the number of schools to help compensate for school non-response. Furthermore, within each state, schools will be implicitly stratified by (1) urbanicity (urban, suburban or rural), (2) school size (small, medium or large), (3) socio-economic status (low, medium or high free/reduced lunch), and (4) race/ethnicity (primarily white, black, Hispanic, other). The second stage will be a random sample of one classroom within each selected school.

- All students within the selected classroom will be tested.
- All five forms will be spiraled within each classroom (randomly distributed)
- A selected school may be tested in only one grade.

Each selected school will randomly select one classroom. This will be done by first alphabetizing the classrooms within the school by teacher's first name for a given grade. The school will select the middle classroom on the list if there are an odd number of classrooms. The school will select the classroom just below the middle if there is an even number of classrooms.

Table 1 shows the number of schools in SBAC sampling frame. The frame is based on the Public Elementary/Secondary Universe Survey Data, which is part of the 2009–2010 Common Core Data (CCD) provided by the National Center for Education Statistics (NCES).

Table 1. Number of Schools in the SBAC Small-Scale Trials Sampling Frame

| State | School Population | | |
|-------|---------|---------|----------|
|       | Grade 4 | Grade 7 | Grade 11 |
| AL | 720 | 471 | 376 |
| CA | 5,642 | 2687 | 2,123 |
| CT | 578 | 285 | 219 |
| DE | 104 | 51 | 44 |
| HI | 201 | 79 | 61 |
| ID | 361 | 198 | 212 |
| IA | 662 | 377 | 394 |
| KS | 714 | 399 | 342 |
| ME | 260 | 154 | 94 |
| MI | 1,762 | 970 | 957 |
| MO | 1,107 | 685 | 561 |
| MT | 313 | 238 | 169 |
| NV | 347 | 130 | 106 |
| NH | 263 | 136 | 88 |
| NC | 1,332 | 634 | 531 |
| ND | 252 | 186 | 162 |
| OR | 719 | 377 | 284 |
| PA | 1,708 | 923 | 685 |
| SC | 607 | 295 | 212 |

| | | | |
|---|---|---|---|
| SD | 277 | 218 | 173 |
| VT | 214 | 123 | 61 |
| WA | 1,102 | 497 | 422 |
| WV | 419 | 200 | 146 |
| WI | 1,105 | 618 | 553 |
| WY | 179 | 91 | 89 |
| Total | 20,948 | 11,022 | 9,064 |

## Explicit Stratification

States will be used for explicit stratification. In explicit stratification the population frame is divided into mutually exclusive strata, and then a sample is drawn within each stratum. In this case the population frame is explicitly stratified by state. Then within each state the population of schools will be implicitly stratified. This is obtained by sorting schools according to the implicit stratification variables, then using systematic sampling to select a simple random sample of schools from the ordered list. The number of schools selected per state is indicated in Table 2. The sampling fraction is the proportion of schools sampled from the population of schools. The sampling fraction is used to sample enough schools to ultimate provide the student sample size required in the SBAC scope of work (about 15,000 students). We assume that about 80% of the selected schools will actually participate.

Table 2. Smarter Balanced School Sample

| | Selected School Sample | | | Target School Sample | | |
|---|---|---|---|---|---|---|
| State | Grade 4 | Grade 7 | Grade 11 | Grade 4 | Grade 7 | Grade 11 |
| AL | 11 | 14 | 11 | 9 | 11 | 9 |
| CA | 85 | 78 | 64 | 68 | 62 | 51 |
| CT | 9 | 8 | 7 | 7 | 6 | 6 |
| DE | 2 | 1 | 1 | 2 | 1 | 1 |
| HI | 3 | 2 | 2 | 2 | 2 | 2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ID | 5 | 6 | 6 | 4 | 5 | 5 |
| IA | 10 | 11 | 12 | 8 | 9 | 10 |
| KS | 11 | 12 | 10 | 9 | 10 | 8 |
| ME | 4 | 4 | 3 | 3 | 3 | 2 |
| MI | 26 | 28 | 29 | 21 | 22 | 23 |
| MO | 17 | 20 | 17 | 14 | 16 | 14 |
| MT | 5 | 7 | 5 | 4 | 6 | 4 |
| NV | 5 | 4 | 3 | 4 | 3 | 2 |
| NH | 4 | 4 | 3 | 3 | 3 | 2 |
| NC | 20 | 18 | 16 | 16 | 14 | 13 |
| ND | 4 | 5 | 5 | 3 | 4 | 4 |
| OR | 11 | 11 | 9 | 9 | 9 | 7 |
| PA | 26 | 27 | 21 | 21 | 22 | 17 |
| SC | 9 | 9 | 6 | 7 | 7 | 5 |
| SD | 4 | 6 | 5 | 3 | 5 | 4 |
| VT | 3 | 4 | 2 | 2 | 3 | 2 |
| WA | 17 | 14 | 13 | 14 | 11 | 10 |
| WV | 6 | 6 | 4 | 5 | 5 | 3 |
| WI | 17 | 18 | 17 | 14 | 14 | 14 |
| WY | 3 | 3 | 3 | 2 | 2 | 2 |
| Total | 317 | 320 | 274 | 254 | 255 | 220 |

|  | | Total > | 911 |  | Total > | 729 |

| Sampling Fractions | | Participation Rate | |
| --- | --- | --- | --- |
| Grade 4 | 0.015 | Grade 4 | 0.80 |
| Grade 7 | 0.029 | Grade 7 | 0.80 |
| Grade 11 | 0.030 | Grade 11 | 0.80 |

## Implicit Stratification

This above sample of schools will be proportionally allocated across the implicit strata. This is referred to as *systematic simple random sampling*. Let's use California to illustrate systematic simple random sampling. In California we have $M$ = 5,642 schools at grade 4 and we want to sample $m$ = 85 schools. In this case we would sort the $M$ schools in California by (1) urbanicity (urban, suburban or rural), (2) school size (small, medium or large), (3) socio-economic status (low, medium or high free/reduced lunch), and (4) race/ethnicity (primarily white, black, Hispanic, other). Then with a random start $k$ between 1 and $M/m$ we systematically sample using a sampling interval $= M/m$. The sorting is done in serpentine order. For example, sort the schools using urbanicity in ascending order. Then within the first level of the urbanicity (urban), sort school size in ascending order. Within the second level of the urbanicity (suburban), sort the school size in descending order. In this way we sort the school size variable to alternate between ascending and descending sorting throughout all levels of the urbanicity. We do the same for free/reduced lunch by sorting the lunch variable within levels formed from the first two variables, again alternating between ascending and descending order. We do the same for the race/ethnicity variable. This sorting algorithm minimizes the change from one school to the next with respect to the sorting variables to make nearby schools more similar. After sorting, we then pick a random starting school and increase by the sampling interval to select the next sampled school. Between different states, we change the sorting order. For example, for the first state, the urbanicity is sorted in ascending order, then for the second state in descending order, etc. Then we do serpentine sorting for the other variables. The number of students to be sampled is shown in Table 3. We assume that about 95% of the selected students will participate.

Table 3. Smarter Balanced Student Sample

| | Selected Student Sample | | | Target Student Sample | | |
|---|---|---|---|---|---|---|
| State | Grade 4 | Grade 7 | Grade 11 | Grade 4 | Grade 7 | Grade 11 |
| AL | 189 | 231 | 225 | 180 | 219 | 214 |
| CA | 1,428 | 1,302 | 1,275 | 1,357 | 1,237 | 1,211 |
| CT | 147 | 126 | 150 | 140 | 120 | 143 |
| DE | 42 | 21 | 25 | 40 | 20 | 24 |
| HI | 42 | 42 | 50 | 40 | 40 | 48 |
| ID | 84 | 105 | 125 | 80 | 100 | 119 |
| IA | 168 | 189 | 250 | 160 | 180 | 238 |
| KS | 189 | 210 | 200 | 180 | 200 | 190 |
| ME | 63 | 63 | 50 | 60 | 60 | 48 |
| MI | 441 | 462 | 575 | 419 | 439 | 546 |
| MO | 294 | 336 | 350 | 279 | 319 | 333 |
| MT | 84 | 126 | 100 | 80 | 120 | 95 |
| NV | 84 | 63 | 50 | 80 | 60 | 48 |
| NH | 63 | 63 | 50 | 60 | 60 | 48 |
| NC | 336 | 294 | 325 | 319 | 279 | 309 |
| ND | 63 | 84 | 100 | 60 | 80 | 95 |
| OR | 189 | 189 | 175 | 180 | 180 | 166 |
| PA | 441 | 462 | 425 | 419 | 439 | 404 |
| SC | 147 | 147 | 125 | 140 | 140 | 119 |
| SD | 63 | 105 | 100 | 60 | 100 | 95 |
| VT | 42 | 63 | 50 | 40 | 60 | 48 |

| | | | | | | |
|---|---|---|---|---|---|---|
| WA | 294 | 231 | 250 | 279 | 219 | 238 |
| WV | 105 | 105 | 75 | 100 | 100 | 71 |
| WI | 294 | 294 | 350 | 279 | 279 | 333 |
| WY | 42 | 42 | 50 | 40 | 40 | 48 |
| Total | 5,334 | 5,355 | 5,500 | 5,071 | 5,090 | 5,231 |
| | | Total > | 16,189 | | Total > | 15,392 |

| Average Class Size | | Participation Rate | |
|---|---|---|---|
| Grade 4 | 21 | Grade 4 | 0.95 |
| Grade 7 | 21 | Grade 7 | 0.95 |
| Grade 11 | 25 | Grade 11 | 0.95 |

Design Effects

In a simple random sample, the information provided by any student within the sample is independent of the information provided by other students in the sample, because there is no dependency in the selection of students. In most educational contexts, however, student samples are clustered, with students within a cluster being more similar than would be expected from a random draw of students from the population. Students within a school tend to be more similar to one another than a random sample of students from across the state would be, due to a variety of factors that pull for similarities among people, including common geographic and socioeconomic factors. This clustering dramatically reduces the efficiency of most samples drawn in educational contexts and greatly increases the sampling error of estimates derived from these samples. The impact of the sampling procedure on the standard errors is referred to as the *design effect*.

Kish (1965) popularized the concept of design effects. Design effects are described in more detail by Cochran (1977), Levy (1999) and Lohr (1999). The traditional measure of the effect of the sample design is how much information is provided from the existing design relative to a sample based on a simple random sample. This is measured by the design effect—the ratio of the variance of a statistic that takes the characteristics of the sample into account over the variance of the same statistic based on a simple random sample of the same size—that is, the same number of individuals selected at random from the entire state population of students in the same grade without regard to school. In the case of sampling schools and then testing all students within the selected schools, as described here, the design effect is always greater than 1, meaning that, in terms of statistical efficiency, such a sample will always provide a larger standard error for any statistic than the standard error that could be provided from a simple random sample of the same size.

AIR recommends a two-stage sample design which first randomly selects schools and then randomly selects a classroom within schools. All students within the selected classroom would be tested. The formula for the two-stage design effect is $Deff = 1 + (\bar{n}_C - 1)\rho_C + \bar{n}_C(\bar{c}_S - 1)\rho_S$, where $\bar{n}_c$ = the average number of students per school per class per form, $\bar{c}_S$ is the average number of classes per school per form, $\rho_C$ is the intra-class correlation within classes, and $\rho_S$ is the intra-class correlation within schools. For example, the design effect for Grade 4 below can be calculated as follows. We assume there will be an average of 21 students per class with one class per school and 5 forms administered within each class. This yields an average of 4.2 students per school per class per form $(\bar{n}_c)$. The average number of classes per school per form is equal to $1(\bar{c}_S)$. Therefore, the design effect is equal to 1+ (4.2-1).15 + 4.2(1-1).10 = 1.48.

The effective sample size is the actual sample size divided by the design effect. This number indicates the size of a simple random sample that would have the same statistical standard error as that produced by the actual stand-alone field test sample when the sample design is taken into account. Table 4 shows the anticipated design effects and effective sample sizes. The class size estimates in Table 3 and Table 4 are based on national estimates provided by NCES (Average Class Size for Public School Teachers in Elementary Schools, Secondary Schools, and Schools with Combined Grades, by Classroom Type [Table 8], U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey [SASS], 2007–2008).

Table 4. Design Effects in the Smarter Balanced Small-Scale Trials

|  | Class Size | Class Intra-class Correlation | School Intra-class Correlation | Number Forms | Target # Responding Schools | Target # Responding Students | Design Effect | Student Sample per Form | Effective Sample per Form |
|---|---|---|---|---|---|---|---|---|---|
| Grade 4 | 21 | 0.15 | 0.10 | 5 | 254 | 5,071 | 1.48 | 1,014 | 685 |
| Grade 7 | 21 | 0.15 | 0.10 | 5 | 255 | 5,090 | 1.48 | 1,018 | 688 |
| Grade 11 | 25 | 0.15 | 0.10 | 5 | 220 | 5,231 | 1.60 | 1,046 | 654 |
|  |  |  |  |  | 729 | 15,392 |  |  |  |

## Power Analysis

Power is the probability that you will be able to detect the effect you are looking for in the data. Power analyses strategies are described in Cohen (1969 & 1988). The effective sample sizes in Table 4 can be used to conduct a power analysis for the small-scale trials. Let's assume that we wish to compare the *p*-values of items under two conditions: hand-scoring $(p_1)$ and automated scoring

$\left( p_2 \right)$. We want to have a sample large enough to have a .80 probability $(1 - \beta)$ of detecting a minimally detectable effect (MDE) of $\delta = .20$. Effect size estimates for *p*-values (with common *n*) can be obtained by $\delta = 2\arcsin\left( \sqrt{p_1} \right) - 2\arcsin\left( \sqrt{p_2} \right)$. Are the planned SBAC sample sizes large enough to meet these criteria if we use a two-tailed *z*-test with $\alpha = .05$? Table 5 shows that the planned sample design would meet and exceed these criteria. For example, for grade 4 we plan to obtain an effective sample size equal to 678 students. This is enough students to detect an effect size of .15, which exceeds our expectation of being able to detect an effect size equal to .20. We obtain similar results for grades 7 and 11.

Table 5. Power Analysis of the Smarter Balanced Small-Scale Trials

| | Minimal Detectable Standardized Effect Size | A Priori Power | Alpha | Number Tails Z-Test | Effective Sample per Form | Actual Detectable Standardized Effect Size |
|---|---|---|---|---|---|---|
| Grade 4 | 0.2 | 0.8 | 0.05 | 2 | 685 | 0.15 |
| Grade 7 | 0.2 | 0.8 | 0.05 | 2 | 688 | 0.15 |
| Grade 11 | 0.2 | 0.8 | 0.05 | 2 | 654 | 0.15 |

*Student Sampling Weights*

*School weight*: The schools will be selected with a systematic simple random sample. If $m_h$ is the number of schools to be selected from stratum *h,* then the probability of the *i*th school within stratum *h* being selected is $p_{hi} = \dfrac{m_h}{M_h}$, where $M_h$ is the total number of schools in stratum *h*. The school weight is $w_{hi} = \dfrac{1}{p_{hi}}$.

*Class weight*: One class will be selected within the selected school, so the selection probability of class *j* in school *i* from stratum *h* is $p_{hij} = 1 / J_{hi}$, where $J_{hi}$ is the number of classes in stratum $h$ and school $i$. The class weight is $w_{hij} = \dfrac{1}{p_{hij}}$.

*Student weight*: Finally, all students are selected from each selected class so that the selection probability of student *k* is $p_{hijk} = 1$. The overall selection probability for student *k* is $p_k = p_{hi} p_{hij} p_{hijk}$.

The overall weight can be calculated as $w_k = \dfrac{1}{p_k}$.

*Normalized weight*: The weight is then normalized within each stratum to the total number of sampled students. More specifically, suppose that there are total of $N_h$ students in stratum *h* and

the weighted sample size for stratum *h* is $W_h = \sum_{k \in h} w_h$. Then, the normalized weight for sampled

student *k* in stratum *h* is $\tilde{w}_k = \dfrac{N_h}{W_h} w_k$.

## Results of School Sampling

At this time AIR has drawn the sample of schools as outlined above. AIR used SAS PROC SURVEY SELECT. The characteristics of the sampling frame compared to the characteristics of the sample are contained in Table 6 – Table 9. Due to rounding in the cells some of the marginal totals involving percentages may not sum to 100%.

In general, the match is very good. This indicates the SBAC sample is representative of the population.

Table 6. Stratification based on Urbanicity

| | | Sampling Frame | | Smarter Balanced Sample | |
|---|---|---|---|---|---|
| | Urbanicity | # Schools | Percent | # Schools | Percent |
| | Urban | 5925 | 28 | 89 | 28 |
| Grade 4 | Suburb | 5686 | 27 | 84 | 26 |
| | Rural | 9337 | 45 | 144 | 45 |
| | Total | 20948 | 100 | 317 | 100 |

| | Urbanicity | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | Urban | 2579 | 23 | 74 | 23 |
| Grade 7 | Suburb | 2231 | 20 | 69 | 22 |
| | Rural | 6212 | 56 | 177 | 55 |
| | Total | 11022 | 100 | 320 | 100 |

| | Urbanicity | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | Urban | 1899 | 21 | 57 | 21 |
| Grade 11 | Suburb | 1807 | 20 | 56 | 20 |
| | Rural | 5358 | 59 | 161 | 59 |
| | Total | 9064 | 100 | 274 | 100 |

Table 7. Stratification based on School Size

| | | Sampling Frame | | Smarter Balanced Sample | |
|---|---|---|---|---|---|
| | School Size | # Schools | Percent | # Schools | Percent |
| | small | 7093 | 34 | 107 | 34 |
| Grade 4 | medium | 6879 | 33 | 106 | 33 |
| | large | 6976 | 33 | 104 | 33 |
| | Total | 20948 | 100 | 317 | 100 |

| | School Size | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | small | 3645 | 33 | 106 | 33 |
| Grade 7 | medium | 3712 | 34 | 109 | 34 |
| | large | 3665 | 33 | 105 | 33 |
| | Total | 11022 | 100 | 320 | 100 |

| | School Size | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | small | 3007 | 33 | 90 | 33 |
| Grade 11 | medium | 3034 | 33 | 93 | 34 |
| | large | 3023 | 33 | 91 | 33 |
| | Total | 9064 | 100 | 274 | 100 |

Table 8. Stratification based on Socio-Economic Status

| | | Sampling Frame | | Smarter Balanced Sample | |
|---|---|---|---|---|---|
| | SES | # Schools | Percent | # Schools | Percent |
| | low | 6983 | 33 | 108 | 34 |
| Grade 4 | medium | 6981 | 33 | 104 | 33 |
| | high | 6984 | 33 | 105 | 33 |
| | Total | 20948 | 100 | 317 | 100 |

| | SES | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | low | 3675 | 33 | 108 | 34 |
| Grade 7 | medium | 3673 | 33 | 105 | 33 |
| | high | 3674 | 33 | 107 | 33 |
| | Total | 11022 | 100 | 320 | 100 |

| | SES | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| | low | 3022 | 33 | 93 | 34 |
| Grade 11 | medium | 3021 | 33 | 87 | 32 |
| | high | 3021 | 33 | 94 | 34 |
| | Total | 9064 | 100 | 274 | 100 |

Table 9. Stratification based on Ethnicity

|  | Sampling Frame | | Smarter Balanced Sample | |
|---|---|---|---|---|
|  | Ethnicity | # Schools | Percent | # Schools | Percent |
| Grade 4 | White | 13963 | 67 | 213 | 67 |
|  | Black | 2118 | 10 | 29 | 9 |
|  | Hispanic | 3874 | 18 | 60 | 19 |
|  | Other | 993 | 5 | 15 | 5 |
|  | Total | 20948 | 100 | 317 | 100 |

| | Ethnicity | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| Grade 7 | White | 7577 | 69 | 222 | 69 |
|  | Black | 1291 | 12 | 37 | 12 |
|  | Hispanic | 1680 | 15 | 51 | 16 |
|  | Other | 474 | 4 | 10 | 3 |
|  | Total | 11022 | 100 | 320 | 100 |

| | Ethnicity | # Schools | Percent | # Schools | Percent |
|---|---|---|---|---|---|
| Grade 11 | White | 6493 | 72 | 195 | 71 |
|  | Black | 916 | 10 | 27 | 10 |
|  | Hispanic | 1331 | 15 | 42 | 15 |
|  | Other | 324 | 4 | 10 | 4 |
|  | Total | 9064 | 100 | 274 | 100 |

# References

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley and Sons.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley and Sons.

Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications* (3rd ed.). New York: John Wiley and Sons.

Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press, Brooks/Cole Publishing.

Appendix B: Small Scale Trials Item Prompts and Scoring Rubrics

| Item Number | Item Description |
|---|---|
| 43173 | The student must explain the process for how both schools can find the least number of tables needed. |
| Prompt and Scoring Rubric for Item 43173 | Prompt: Describe a strategy using words that both schools could use to determine the least number of tables required.<br><br>Type your answer in the space provided.<br><br>2 Point Text: The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes both of the correct strategies:<br><br>1st Strategy<br>• Divide the number of students at each school by 6.<br>OR<br>• Repeated subtraction of 6 from the number of students.<br><br>2nd Strategy<br><br>• Since the number of tables purchased must be a whole number, round to the next whole number to ensure that all students have a place to sit.<br><br>1 Point Text: The response addresses the task in a partially satisfactory manner. It is partially complete, containing enough information (general or specific) to answer part of the question.<br><br>For this item, the response includes one of the following correct strategies:<br>• Divide the number of students at each school by 6.<br>• Repeated subtraction of 6 from the number of students.<br>• Since the number of tables purchased must be a whole number, round to the next whole number to ensure that all students have a place to sit.<br>• Use 13 tables and 6 tables.<br>• Use a total of 19 tables.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly no other information is shown. The student may have written on a different topic, or written "I don't know." |
| 43248 | The student must explain how the author defines "art" and support the definition with a detail from the text. |

| | |
|---|---|
| Prompt and Scoring Rubric for Item 43248 | Prompt: How would the author of this passage define art? Use details from the passage to support your answer.<br><br>Correct Responses: The author would probably think that anything can count as art depending on the individual's view.  This is shown in the text where Clara sees a tablecloth and a spider web as art.<br><br>2 Point Text: The response includes at least one statement from each category:<br>Author's Definition:<br>   ⓪ Art can be viewed in many different ways.<br>   ⓪ Art is interpreted by the individual / interpreted differently.<br>   ⓪ Many things count as art / art is in the eye of the beholder.<br>   ⓪ Everyday/ ordinary objects can be art.<br>   ⓪ Anything with complexity / craftsmanship can be art.<br>Details:<br>   ⓪ Clara views geometry as art / there is a relationship between geometry and art.<br>   ⓪ Clara views the tablecloth as art.<br>   ⓪ Clara sees the spider web as art / calls the spider an artist.<br><br>1 Point Text:  The response includes at least one statement from only one of the above categories.<br>Sample answer: The author would probably think that art can be viewed in many different ways.<br><br>0 Point Text:  The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know."<br>Sample answer: N/A |
| 43280 | The student must explain why a game's rules are important to the characters in both texts. A full-credit answer must reference both texts. |
| Prompt and Scoring Rubric for Item 43280 | Prompt: Student Directions for Parts 1 and 2<br>Part 1 (35 Minutes)<br><br>Your task: You will read four sources: two articles and two stories.  Then you will answer a question about what you learned.  In Part 2, you will write a story about someone involved in or watching a game, event, or sport.<br><br>Steps to Follow: In order to plan and write your story, you will do all of the following:<br>   1) Examine several sources<br>   2) Answer a question about sources<br><br>Directions for Beginning: You will now examine several sources.  You can re-examine any of the sources as often as you like.<br><br>Research Question: |

After examining the sources, use the remaining time in Part 1 to answer a question about them. Your answer to the question will be scored. Also, your answer will help you think about the sources you have read and viewed, which should help you write your story.

The authors of Documents #3 <u>The Invention of Kickball</u> and #4 <u>Casey's First Match</u> write stories about games. The rules of each game are a central idea in both stories. How are the rules of each game important to Jacob in the story of the invention of kickball (Document #3) and to Casey in the story of the math competition (Document #4)?

Use details from each story to support your answer.

Correct Answer: The rules are important to Jacob in his story because he needs to explain them to his uncle. The rules are important to Casey in her story because she wants to win a competition.

<u>2 Point Text:</u>
Response includes a correct way the rules of kickball are important to Jacob with a correct way the rules of a mathlete competition are important to Casey.

Jacob:
- He needs to explain them.
- He tells them to his uncle.
- He shows the game to his uncle.
- He and his friends invented the game.
- He uses them to help his uncle.

Casey:
- It is her first time in a Math/Mathlete competiton.
- She doesn't know the rules of the Mathlete competition.
- If she doesn't know the rules, she wouldn't know how to play.
- She is on a Mathelete team.
- She wants to win/help her team.
- She has to solve Math problems in a competiton.
- She sits on the stage/waits for her turn to solve Math problems.

<u>1 Point Text:</u>
Response includes a correct way the rules of kickball are important to Jacob but a missing or incorrect way the rules of a Mathlete competition are important to Casey.
OR
Response includes a correct way the rules of a mathlete competition are important to Casey but a missing or incorrect way the rules of kickball are important to Jacob.
Sample Answer: The rules are important to Jacob in his story because he needs to explain them to his uncle.

<u>0 Point Text:</u>
The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only

| | |
|---|---|
| | repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know."<br>Sample Answer: N/A |
| 43297 | The student must explain how the structure of the text sets up the ending. A specific detail illustrating the setup or resolution is required for full credit. |
| Prompt and Scoring Rubric for Item 43297 | Prompt: How does the structure of this text prepare the reader for the ending? Support your response with evidence from the text.<br><br>Type your answer in the space provided.<br><br>Correct Responses:  The structure prepares the reader for the ending using dialogue between the pieces of the clock to show why the clock stopped in the first place.  This sets up the ending where the farmer thinks his watch has gained 30 minutes overnight because it doesn't match the clock.<br><br>2 Point Text:<br>Response includes one correct explanation about how the structure of the text prepares the reader for the ending with a supporting detail.<br><br>**How the structure prepares the reader for the ending:**<br>• The text includes dialogue/the parts of the clock argue with each other.<br>• The text uses cause and effect to explain the problem.<br>• The reader learns why the clock has stopped.<br>• The pacing of the story suggests that the problem will be solved by the end.<br>• The text sets up a conflict and a resolution.<br>• The text uses foreshadowing.<br><br>**Support:**<br>• The clock stops because the pendulum is tired of ticking.<br>• The pendulum is the only part that wants to stop working.<br>• "the pendulum, who spoke thus: 'I confess myself to be the sole cause of the present stoppage'"<br>• The dial plate/other pieces of the clock urge the pendulum to work.<br>• The pieces tell the pendulum how important he is to the clock.<br>• "You have done a great deal of work in your time; so have we all, and are likely to do; which, although it may fatigue us to think of, the question is, whether it will fatigue us to do"<br>• "Recollect that, although you may think of a million strokes in an instant, you are required to execute but one; and that, however often you may hereafter have to swing, a moment will always be given you to swing in."<br>• The farmer's watch gains half an hour.<br>• The farmer thinks his watch is wrong.<br><br>1 Point Text:<br>Response includes one correct explanation about how the structure of the text prepares the reader for the ending with a missing or incorrect supporting detail. |

| | | |
|---|---|---|
| | Note: The student will NOT receive credit for a correct supporting detail without a correct structure. Sample Answer: The structure tells the reader what the lesson is going to be because the clock parts are arguing about it.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know."<br>Sample Answer: N/A | |
| 43559 | The student must choose the "best" proof of the Pythagorean Theorem and provide support for the decision. | |
| Prompt and Scoring Rubric for Item 43559 | Prompt: Why is Attempt 2 the best proof?<br><br>Type your answer in the space provided.<br><br>Correct responses: Attempt number 2 works for any size of right triangle.<br>Attempt number 1 only shows that the theorem works for a 3, 4, 5 triangle.<br>Attempt number 3 only shows that the theorem works for isosceles right triangles.<br><br>2 Point Text:<br>N/A<br><br>1 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br>• Attempt 2 works for any right triangle/algebraically.<br>☇ Attempt 1 only works for a 3, 4, 5 triangle and Attempt 3 only works for isosceles right triangles.<br>☇ Attempt 2 works for any right triangle/algebraically. Attempt 1 only works for a 3, 4, 5 triangle and Attempt 3 only works for isosceles right triangles.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." | |
| 43397 | The student must identify a technique the author uses and provide an example from the text. | |
| Prompt and | Prompt: What techniques does the author use to convince the reader of the believability of the events in the text? | |

| | |
|---|---|
| **Scoring Rubric for Item 43397** | Provide details from the text to support your answer.<br><br>Correct Responses: The author uses a first-person narrator to tell us the facts about everything that happened and to let us know how he feels about Conway and Barting.  In the passage, it states that Foley trusts Barting because he was an "honorable" and "truthful" man with whom he had served in a war, so the reader can believe that the narrator is telling the truth about his story.<br><br>2 Point Text:<br>The response includes a correct technique and a correct supporting detail from the text.<br><br>**Techniques:**<br>• Straightforward/casual/conversational tone<br>• First-person narrative<br>• Descriptions/characterization of Conway/Barting<br>• Details about Conway and Barting's friendship<br>• Realistic setting/details about the time period<br>• Foreshadowing<br><br>**Support:**<br>• "This is a story told by the late Benson Foley of San Francisco" (only support for first person)<br>• "In the summer of 1881 I met a man named James H. Conway, a resident of Franklin, Tennessee."<br>• "I had known Barting as a captain in the Federal army during the civil war."<br>• "Barting had always seemed to me an honorable and truthful man."<br>• "the warm friendship which he expressed in his note for Mr. Conway was to me sufficient evidence that the latter was in every way worthy of my confidence and esteem."<br>• "I had in my pocket a photograph of Barting . . . without a mustache."<br>• "it had been solemnly agreed between him and Barting that the one who died first should, if possible, communicate with the other" (only support for foreshadowing)<br><br>1 Point Text:<br>The response includes one of the correct techniques without a correct supporting detail from the text. Note: The student will NOT receive credit for a correct supporting detail without a correct technique.<br>Sample Answer:  The author makes the story more believable by telling it in the first-person perspective, like we are listening to the narrator tell us about his own story.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43403 | The student must write an ending for an unfinished story. Specific details and appropriate word choice/vocabulary are required for full credit. |

| | |
|---|---|
| Prompt and Scoring Rubric for Item 43403 | Prompt: *The following two paragraphs are the beginning of a story about trying to have fun on a rainy day. Read the paragraphs. After reading the paragraphs, you will finish the story.*<br><br>Last Saturday morning, I woke with my alarm at 8 a.m. and leapt out of bed. My softball team was having its very first game that day, and I couldn't wait. I had been practicing for weeks, and I knew I was ready. I hurried across my bedroom to where my uniform was carefully folded and placed on a chair. It was only then that I looked out the window. It was pouring rain! The game would be canceled for sure; I felt ready to cry because I was so upset.<br><br>Then, I heard my mom and brother laughing downstairs. The sound made me forget my disappointment. Curious, I wandered out of my bedroom, still in my pajamas, and made my way down the stairs.<br><br>Now finish the story.<br><br>Type your answer in the space provided.<br><br>Correct Responses: I followed the laughter into the living room. My mom and my brother sat on the floor, flipping through the book of old family photographs. I wondered what could possibly be so funny. Then, my brother looked up and saw me standing there.<br><br>"Come here, you have to see these!" he said. I glanced out the window; the rain was still pouring down in sheets. I thought that if I couldn't play softball today, I might as well try to have fun inside. I sat down next to my brother, and he showed me pictures of us as babies and toddlers, playing games and making silly faces for the camera. In a few minutes, I was laughing too. Maybe spending the day indoors wouldn't be so bad.<br><br><u>2 Point Text:</u><br>The response:<br>　⓾　provides appropriate and predominately specific details or evidence<br>　⓾　uses appropriate word choices for intended audience and purpose<br><br><u>1 Point Text:</u><br>The response:<br>　⓾　provides mostly general details and evidence, but may include extraneous or loosely related details<br>　⓾　has a limited and predictable vocabulary that may not be consistently appropriate for the intended audience and purpose<br>Sample Answer: I followed the sounds into the living room. My mom and my brother sat on the floor, flipping through the book of old family photographs. What could possibly be so funny?<br><br>"Come here, you have to see these!" he said. I glanced out the window; the rain was still pouring down in sheets. I thought that if I couldn't play softball today, I might as well try to have fun inside. I sat down next to my brother, and he showed me pictures. In a few minutes, I was laughing too. Maybe spending the day indoors wouldn't be so bad. |

| | |
|---|---|
| | <u>0 Point Text:</u><br>The response:<br>    ⦿  includes few supporting details that may be vague, repetitive, incorrect, or interfere with the meaning of the text<br>    ⦿  has inappropriate vocabulary for the intended audience and purpose |
| 43412 | The student must identify a characteristic of a successful mail delivery service and provide a supporting detail from the text. |
| Prompt and Scoring Rubric for item 43412 | Prompt: What does the passage suggest is important for a successful mail delivery service? Support your answer with a detail from the passage.<br><br>Type your answer in the space provided.<br><br>Correct Responses: People were looking for a mail delivery service that was always fast. The Pony Express did this by sending riders from Missouri to California in just eleven days.<br><br><u>2 Point Text:</u><br>What was important in a mail delivery service:<br>    ⦿  fast / speed<br>    ⦿  reliable / reliability / consistency / being on time<br>    ⦿  trustworthy / trust<br><br>Supporting detail:<br>    ⦿  Stations were built ten to fifteen miles apart / riders would change horses and then continue on.<br>    ⦿  They promised mail would travel across the country in ten days.<br>    ⦿  Riders rode from California to Missouri in eleven days.<br>    ⦿  Riders traveled through bad weather.<br>    ⦿  "No matter what … they always delivered the mail."<br>    ⦿  People switched to the telegraph because it was faster.<br><u>1 Point Text:</u><br>The response contains an example of what was important for a mail delivery service.<br>OR<br>The response contains a correct supporting detail.<br>Sample Answer: People wanted a mail delivery service that they could trust all of the time.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43416 | The student must explain how a character feels at the end of the text and support this feeling with a detail from the text. |

| | |
|---|---|
| Prompt and Scoring Rubric for item 43416 | Prompt: Explain how Alexa most likely feels when her sister Marisa plays piano. Use information from the text to support your explanation.<br><br>Correct Responses: Alexa feels proud of her sister. She helped Marisa practice and gave her advice, so she wants to see her do well.<br><br><u>2 Point Text:</u><br>Response includes a correct feeling and a correct explanation.<br><br>Feeling:<br>    • proud / pride<br>    • excited / excitement<br>    • happy / happiness<br>    • calm / patient<br>    • confident<br><br>Explanation:<br>    • She talks Marisa into joining the talent show.<br>    • She helps Marisa practice.<br>    • She gives Marisa advice.<br>    • She tells Marisa a trick to be less nervous.<br>    • She encourages Marisa.<br>    • She tells Marisa "that was really good" / "you'll be great."<br>    • She tells Marisa not to worry / says "don't worry" / "you'll be fine."<br>    • She smiles up at Marisa.<br><br>Note: Vague feelings such as "good," "great," or "fine" will NOT receive credit.<br><br><u>1 Point Text:</u><br>Response includes a correct feeling without a correct explanation.<br><br>Note: The student will NOT receive credit for a correct explanation without a correct feeling.<br>Sample Answer: Alexa is excited to see her sister finally play.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43422 | The student must explain how the author uses an example to support his/her main idea. For full credit, the student must identify a different but analogous example from the text. |
| Prompt and Scoring | Prompt: Read this sentence from the text.<br><br>"The idea is that when people smell the cookies, they will feel good, and as a result they will want to buy |

| | |
|---|---|
| Rubric for item 43422 | the house." |

Explain how this sentence relates to the author's main point. Then, give another example from the passage that relates to the author's point in a similar way.

Type your answer in the space provided.

Correct Responses: Smell is connected to people's feelings and can make them spend money on things. The smell of chlorine brings up pool-related memories.

2 Point Text:
The response relates the sentence to the main idea and adequately supports the reasoning with another example. Correct answers include:

Relates to Main Idea:
- ✪ The smell connects to memory / emotion / feelings.
  - ✪ The smell can make you feel good.
  - ✪ The smell makes you feel things.
  - ✪ The smell makes you feel a certain way.
AND
- ✪ The smell gets people to buy things/spend money / act a certain way.
  - ✪ The smell influences people / decisions.

Note: Students must explain that smell is both linked to memory / emotions AND that smell influences how people act.

Another Example:
- ✪ Chlorine makes people think of pool-related memories.
- ✪ Babies like garlic if they're introduced to it before birth.
- ✪ Grocery stores smell like baked goods.
- ✪ Shoe companies make their shops smell good.
- ✪ Malodorants repel people.
- ✪ Police use stink bombs to make people leave / break up riots.
- ✪ Skunks use smell as a defense.

1 Point Text:
The response tells how the sentence relates to the main idea but does not give another example.

Note: Students will NOT receive credit for a correct example without a correct relation.
Sample Answer: Smell and emotions are related because smells can influence your feelings and decisions.

0 Point Text:
The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is

| | |
|---|---|
| | shown. The student may have written on a different topic or written, "I don't know." |
| 43435 | The student must identify the attitude of persons named in the text and provide evidence of that attitude. |
| Prompt and Scoring Rubric for Item 43435 | Prompt: Explain what the text suggests about the McWhirters' attitude toward the public. Provide evidence from the text that supports this inference.<br><br>Type your answer in the space provided.<br><br>Correct Responses: The McWhirters show concern for the public by removing records that encourage unsafe or immoral behavior.<br><br><u>2 Point Text:</u><br>Response includes a correct attitude with correct evidence.<br><br>Attitude:<br>● They respected / served / were concerned / cared about the public / understood what the public wanted and supplied it.<br>● They were interested in informing / educating the public with facts.<br><br>Evidence:<br>● They "scoured the globe" / worked hard to collect their data.<br>● They responded to public demand / thirst for unusual knowledge.<br>● They sought out "increasingly obscure, little-known facts."<br>　● They included bizarre / strange / weird / wacky records.<br>● Their website covers facts for a wide range of interests.<br>● They invited the public to submit applications for new records.<br>● Records have been removed for ethical / moral / safety reasons.<br><br><u>1 Point Text:</u><br>Response includes a correct attitude without evidence.<br><br>Note: The student will NOT receive credit for correct evidence without a correct attitude.<br>Sample Answer: The McWhirters wanted to serve the public. That's why they put out their book.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43445 | The student must explain how the author's use of chronological order helps provide characterization and support the answer with an example from the text. |

| | |
|---|---|
| | |
| Prompt and Scoring Rubric for Item 43445 | Prompt: How does the author's use of chronological order in the text highlight Sojourner Truth's personal characteristics?<br><br>Use examples from the text to support your answer.<br><br>Type your answer in the space provided.<br><br>Correct Responses: The author shows how determined Sojourner Truth was over the years of her life. This is particularly evident when the author mentions that even though she didn't have a formal education, she kept standing up and fighting for what was right.<br><br><u>2 Point Text:</u><br>Response includes the following correct use of chronological order to highlight Sojourner Truth's personal characteristics with correct support.<br><br>Use:<br><ul><li>It shows how her life and background influenced her.</li><li>It shows that she was determined / strong / brave / hardworking throughout her life.<ul><li>It shows that she never gave up.</li></ul></li><li>It shows the challenges she faced and overcame during her life.</li><li>It shows her growth/change over time.</li></ul><br>Support:<br><ul><li>Lack of education did not stop her.</li><li>She continued working into her eighties.</li><li>She went from being a slave to inspiring many people.</li><li>She was the first former slave to win a court case.</li><li>She was one of the African American pioneers.</li></ul><br><u>1 Point Text:</u><br>Response includes a correct use of chronological order to highlight Sojourner Truth's personal characteristics without a correct support.<br>Note: The student will NOT receive credit for a correct support without a correct use.<br>Sample Answer: The author uses chronological order to show how determined Sojourner Truth was through her entire life.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43446 | The student must write a brief argumentative essay for or against public transportation. A full-credit response must include evidence from a provided table and use appropriate vocabulary. |

| | |
|---|---|
| Prompt and Scoring Rubric for item 43446 | Prompt: Joseph's English teacher assigned each student the task of writing an argumentative essay requiring a research component. The following table contains Joseph's notes from his research about the merits of using public transit versus driving a private vehicle. Read his notes. Then compose a brief argumentative essay. |

| Public Transit | | Private Vehicle | |
|---|---|---|---|
| Pros: | Cons: | Pros: | Cons: |
| Saves money in gas, car maintenance, and repairs | Have to sacrifice personal space | Is convenient because car is only steps away | May be difficult to find parking |
| Saves frustration by avoiding heavy traffic | Can be difficult to follow schedule | Is more comfortable than public transit | Have to deal with heavy traffic |
| Is better for the environment | Can be unreliable | Can go anywhere and can get there directly | Can only use time in the vehicle for driving |
| Can use time on transit productively | Lacks the privacy and comfort of a car | Is not limited to areas that public transit serves | Costs more than public transportation |
| | May need to tolerate inconsiderate people | | Uses more natural resources |

Compose a brief argumentative essay either in support of, or in opposition to, using public transportation. Use evidence from the table to support your answer.

Type your answer in the space provided.

Correct Responses: Driving a private car is far superior to the use of public transportation. In many situations, it does not make any sense to take public transit, especially when outside of a big city. Far more people in the country use private cars instead of public transportation. This is likely because cars are more convenient for most citizens. For instance, while one's car may only be a few steps away, many residents live a mile or more from the nearest bus stop. The inconvenience of distance is only made worse by the unreliable nature of public transit in general, and buses in particular.

 Furthermore, while some may argue that cars are not worth the extra cost, the time that one saves by driving more than makes up for the increased cost. As they say, time is money. Similarly, without a car one is limited to the geographical area that public transit serves, whereas a private car allows the driver the freedom to travel anywhere he or she desires. And you can't put a price on freedom.

 Though public transit does have its advantages, it is mainly useful for those in heavily serviced areas. For the rest of us, cars are the clear choice for their convenience and the freedom they allow.

2 Point Text:

| | |
|---|---|
| | Evidence/Elaboration: |
| | The response: |
| | • provides appropriate and predominantly specific details or evidence |
| | • uses appropriate word choices for intended audience and purpose |
| | 1 Point Text: |
| | Evidence/Elaboration: |
| | The response: |
| | • provides mostly general details and evidence, but may include extraneous or loosely related details |
| | • has a limited and predictable vocabulary that may not be consistently appropriate for the intended audience and purpose |
| | Sample Answer: There are tons of bus stops near my house. However, I like to use the ones that I can walk to because it makes it easier to catch the bus to school. I know all of the bus schedules so for me the buses are easy to figure out. There's even an application for mobile devices that lets you see when the next bus is coming. This can be really useful when the buses are being unreliable like they can be. |
| | I always use the bus which means that I choose public transportation. Even if I could have a car I wouldn't want it. Cars are hard to drive and can be really dangerous. I'd bike instead of driving if I had the choice. |
| | I like the fact that I get to use the bus and train every day. It makes my trip to school interesting and enjoyable. Sometimes I even read there. I couldn't do that in a car. I think everyone should use public transportation because it's so much better than driving. |
| | 0 Point Text: |
| | Evidence/Elaboration: |
| | The response: |
| | • includes few supporting details that may be vague, repetitive, incorrect, or interfere with the meaning of the text |
| | • has inappropriate vocabulary for the intended audience and purpose |
| | Sample Answer: Public transportation is the best. |
| 43468 | The student must use the provided documents to identify three scientific inaccuracies in a science fiction story. |
| Prompt and Scoring Rubric for item 43468 | Prompt: Part 1 (35 Minutes) You will read several documents about science and science fiction. Then you will answer a question about what you have read. In Part 2, you will write a narrative story about living on the moon or traveling to the moon. You will use current scientific knowledge to shape the story. Steps to Follow |

In order to plan and write your narrative story, you will do the following:

1. Examine several documents.

2. Answer a question about the documents.

Directions for Beginning

You will now examine several documents. You can re-read the documents as often as you like.

Research Question

After examining the research documents, use the remaining time in Part 1 to answer a question. Your answer to this question will be scored. Also your answer will help you think about the research documents you have read, which will help you write your narrative story.

Read all the documents. When you are finished reading the documents provided, review the story, Document #2: Lost on the Moon. This story was written many years ago and used scientific information about the Moon during that time. There are many misunderstandings about the Moon in this science fiction story because we have newer scientific information today. Explain at least three points in Lost on the Moon that are incorrect based on today's scientific information provided in the other documents about the Moon. Use information from the other documents provided to support your answer.

Type your answer in the space provided.

2 Point Text:
A correct response includes at least three of the following points about the Moon that are incorrect based on today's scientific information:

- ⑩ speed of travel to the Moon
- ⑩ amount of gravity
- ⑩ ability to grow food / plants / vegetation
- ⑩ Moon supporting life / people on the Moon
- ⑩ diamonds on the Moon
- ⑩ atmosphere / oxygen on the Moon
- ⑩ water on the Moon

1 Point Text:
Response includes only two points about the Moon that are incorrect based on today's scientific information.

Note: The student will NOT receive credit for a response that includes only one point about the Moon that is incorrect based on today's scientific information.

0 Point Text:
The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only

| | |
|---|---|
| | repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43491 | The student must explain how one document is different from the others. The student must also identify evidence the author uses to strengthen his claim. |
| Prompt and Scoring Rubric for item 43491 | Prompt: Explain how the editorial "For Love of the City" is different from the other sources. Then, explain how the author strengthens his claim that "cities are beneficial for living."<br><br>Type your answer in the space provided.<br><br><u>2 Point Text:</u><br>A full-credit response includes a correct explanation with a correct technique.<br><br>How the editorial is different:<br><ul><li>The author is against rural living / suggests problems with rural living.</li><li>The author portrays rural living as inferior to living in the city.</li><li>The text is a persuasive / opinion piece.<ul><li>The text is a letter to an editor.</li><li>The author gives his own opinions.</li></ul></li></ul>How author strengthens his claim:<br><ul><li>presents points and counterpoints<ul><li>disproves assumptions</li></ul></li><li>cites research / experts</li><li>links cities to human nature<ul><li>humans need interaction with other humans</li></ul></li><li>links cities to historical achievements / progress<ul><li>cities are a source of innovation</li></ul></li><li>implies environmental / moral / ethical / responsibilities and concerns<ul><li>appeals to the reader's sense of responsibility</li><li>rural living uses more fuel / electricity / resources</li><li>rural living creates more pollution</li><li>city living saves money</li></ul></li></ul><u>1 Point Text:</u><br>Response includes only either an explanation of how the editorial is different or how the author strengthens his claim.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |

| | |
|---|---|
| 43497 | The student must write a conclusion to an argumentative essay while acknowledging listed counter-arguments. A full-credit response includes specific details and appropriate word choice. |
| Prompt and Scoring Rubric for item 43497 | **Prompt:** *A local newspaper has written articles on restaurants in your neighborhood. Your teacher has asked you to write an argumentative essay on a topic concerning restaurants, using the articles as sources. After reading the sentences, you will finish the paragraph by including the counter-arguments below. The following is the beginning of your argumentative essay considering whether restaurants should offer more vegetarian-friendly options. After reading the sentences, finish the paragraph by including the counter-arguments below.*<br><br>    I think that restaurants should offer more options for people who prefer a vegetarian diet. Many people these days choose to stay away from meat products. A recent survey found that 38% of people are vegetarians. However, there are few items on restaurant menus for vegetarians to choose from, and the few options that are available are often much more costly than the meat-based alternatives.<br><br>Now finish the argumentative essay while acknowledging these counter-arguments:<br><br>Counter-Arguments:<br>   ⑩   Vegetarian products can be expensive.<br>   ⑩   Almost every restaurant offers salads for vegetarians.<br>   ⑩   There are more people who eat meat than there are vegetarians.<br><br>Type your answer in the space provided.<br><br>**Correct Responses:**<br>I think that restaurants should offer more options for people who prefer a vegetarian diet. Many people these days choose to stay away from meat products. However, there are few items on restaurant menus for vegetarians to choose from. Most restaurants offer salads but people can get tired of eating a salad every day. And even though some vegetarian options can be expensive, I think that if restaurants offered these products, more vegetarians would eat there!<br><br>2 Point Text:<br>The response:<br>• provides appropriate and predominately specific details or evidence<br>• uses appropriate word choices for intended audience and purpose<br><br>1 Point Text:<br>The response:<br>• provides mostly general details and evidence, but may include extraneous or loosely related details<br>• has a limited and predictable vocabulary that may not be consistently appropriate for the intended audience and purpose<br>Sample Answer: I think that restaurants should offer more options for people who prefer a vegetarian diet. There are few items on restaurant menus for vegetarians to choose from. Most restaurants say that they offer salads and that vegetarian options can be expensive. I think that if restaurants offered other |

| | |
|---|---|
| | products, more vegetarians would eat there!<br><br>0 Point Text:<br>The response:<br> •includes few supporting details that may be vague, repetitive, incorrect, or interfere with the meaning of the text<br> •has inappropriate vocabulary for the intended audience and purpose |
| 43546 | The student must explain the process for determining the composition of the crown. |
| Prompt and Scoring Rubric for item 43546 | Prompt: Explain in words the process you could use to determine whether the crown is all gold or a mix of gold and silver.<br><br>Type your answer in the space provided.<br><br>2 Point Text:<br>N/A<br><br>1 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br>· I would first figure out the volume of a pure gold mass that is equal to the King's crown's mass and then compare the two masses. Then I would set up two equations with two unknowns and solve. The unknowns would be the amount of gold and the amount of silver in the crown.<br>· I would set up a proportion to find out the mass of pure gold then solve the proportion.<br>· Find volume of 1.8 kg of pure gold. Compare to volume of crown.<br>· Find mass of 125 cm^3 of pure gold. Compare to mass of crown.<br>· Create a proportion with the crown's mass and volume compared to silver's mass and volume. Create a proportion with the crown's mass and volume compared to gold's mass and volume.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43551 | The student must provide support from the stem for why card A is the best card for the game. |
| Prompt and Scoring Rubric for item | Prompt: Explain why Bingo Card A is the best card to use for this game.<br><br>Type your answer in the space provided.<br><br>2 Point Text: |

| 43551 | N/A |
|---|---|
| | **1 Point Text:**<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br> &bull; Card A contains middle numbers because they have more combinations.<br> &bull; Card A contains more likely numbers because they have more combinations.<br> &bull; Card A contains more likely numbers because they have more different ways.<br> &bull; Card A contains more likely sums.<br> &bull; Card B contains 1 and 17, which are impossible, so this card cannot win. Card C contains extreme and unlikely/less likely numbers because they have few combinations.<br> &bull; Card B contains 1 and 17, which are impossible, so this card cannot win. Card C has higher and lower numbers.<br> &bull; Card B contains 1 and 17, which are impossible, so this card cannot win. Card C has 16 which is the highest/hardest.<br> &bull; Card B contains 1 and 17, which are impossible, so this card cannot win. Card C has 2 which is the lowest/hardest.<br><br>**0 Point Text:**<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43552 | The student must explain the process for how to find the new radius. |
| Prompt and Scoring Rubric for item 43552 | Prompt: Use either words or an equation to tell or show the steps you took to determine the new radius.<br><br>**2 Point Text:**<br>N/A<br><br>**1 Point Text:**<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct strategies:<br><br> &bull; Find the volume of the standard tank. Set the combined volume formulas of cylinder and sphere equal to twice the volume of the standard tank, which gives you the equation: $10r^2 + 4\frac{r^3}{3} = 252$. Then you can use trial and error to determine which radius makes this equation true.<br><br>**0 Point Text:** |

| | The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
|---|---|
| 43555 | The student must explain the process for how to find the lowest cost. |
| Prompt and Scoring Rubric for item 43555 | Prompt: Explain a strategy Max could use to achieve the lowest possible cost.<br><br>Type your answer in the space provided.<br><br>Correct responses: Increase the number of large taxis<br>and<br>decrease the number of empty seats<br>Making sure that there are no empty seats will allow Max to save money. If there are 9 large taxis and 3 small taxis, it will cost $687<br><br>2 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of both of the correct strategies:<br><br>1st Strategy:<br><br>  ✍  Increase the number of large taxis.<br>OR<br>  ✍  Repeated subtraction of 7.<br><br>2$^{nd}$ Strategy:<br><br>  ✍  Use 3 small taxis or minimize the amount of empty seats in taxi cabs.<br><br>1 Point Text:<br>The response addresses the task in a partially satisfactory manner.  It is partially complete, containing enough information (general or specific) to answer part of the question.<br>For this item, the response includes one of the correct strategies:<br><br>· Increase the number of large taxis to 9.<br>· Repeated subtraction of 7.<br>· Use 3 small taxis or minimize the amount of empty seats in taxi cabs.<br>· Use 3 small taxis and 9 large taxis. (No further explanation given)<br>· Use any amount of small and large taxis that has at least 75 total seats and has more large than small taxis.<br>· Find unit rate of both taxis and compare.<br>Sample Answer: Using as few taxis as possible will be the cheapest. Use all large taxis. |

| | |
|---|---|
| | <u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43557 | The student must state if Anne's method is appropriate for the situation and provide support for the decision. |
| Prompt and Scoring Rubric for item 43557 | Prompt: Would it make sense to use Anne's method when determining the temperature to decide what to wear outside? Why or why not?<br><br>Correct Responses: Yes, Anne's method is appropriate for determining what to wear outside based on the weather. The exact answer is not needed because what you wear is not based on an exact number.<br><br><u>2 Point Text:</u><br>N/A<br><br><u>1 Point Text:</u><br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br>· Yes, an exact answer is not necessary.<br>· Yes, temperature is within a few degrees/only a few degrees off.<br>· Yes, estimate is close to the actual answer.<br>· Yes, an approximate answer is close to the actual answer.<br>· Yes, Anne's method is close enough.<br>· Yes, the temperature is close enough.<br>· Yes, the temperature is within a close range.<br>· Yes, it is ok to not have an accurate temperature.<br><br>Note: "Yes, it is close enough for most purposes" (repeating the stem) does not receive credit.<br>Sample Answer: Yes, Anne's method is close enough.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |

| 43564 | The student must give the dimensions of a rectangle with a perimeter of 18. |
|---|---|
| Scoring Rubric for item 43564 | Prompt: SPACE 2<br><br>2 Point Text:<br>N/A<br><br>1 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br> ✎ The rectangle has sides of 6 feet and 3 feet.<br> ✎ A rectangle with sides 6 feet and 3 feet would have an area of 18 square feet and a perimeter of 18 feet.<br> ✎ The rectangle has a length of 6 feet and a width of 3 feet.<br> ✎ The rectangle has a width of 6 feet and a length of 3 feet.<br> ✎ The rectangle has dimensions of 6 feet and 3 feet.<br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know."<br>Sample Answer: It is not possible to create an 18 square foot rectangle with an 18 foot perimeter. |
| 43572 | The student must provide support for why the line is a line of symmetry. |
| Prompt and Scoring Rubric for item 43572 | Prompt: Explain why the line you drew is a line of symmetry.<br><br>Type your answer in the space provided.<br><br>2 Point Text:<br>N/A<br><br>1 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br>• The sides are the same.<br>• The edges will match.<br>• The line is down the middle and there is only one way to go down the middle.<br>• The line is down the middle and it is creates equal parts.<br>• It is folded into matching parts. |

| | |
|---|---|
| | • It will fold evenly together.<br>• It will fold into the same parts.<br>• When you fold the leaf in half, the parts are the same.<br>• It is in halves.<br>• The leaf is divided equally in half.<br>• It divides the leaf into the same/equal parts.<br><br>0 Point Text:<br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43639 | The student must state if Anne's method is appropriate for the situation and provide support for the decision. |
| Prompt and Scoring Rubric for item 43639 | Prompt: Would it make sense to use Anne's method when determining the temperature of an oven? Why or why not?<br><br>Correct Responses: No, in cooking an exact answer is necessary in order to follow a recipe.  Anne's method is close enough to the exact temperature of an oven.<br><br>2 Point Text:<br>N/A<br><br>1 Point Text:<br>The response addresses the task in a satisfactory manner. It is complete and accurate, containing enough information (general or specific) to answer the question thoroughly.<br><br>For this item, the response includes one of the correct explanations:<br><br>· No, in cooking an exact answer is necessary in order to follow a recipe.<br>· No, Anne's method is not close enough to the exact temperature of an oven.<br>· No, an approximation is not appropriate.<br>· No, an estimate is not appropriate.<br>· No, in cooking a right temperature is necessary.<br>· No, the temperature cannot be a little over or under.<br>· No, Anne's method is not accurate to find the temperature of an oven.<br><br>0 Point Text:<br><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |

| | |
|---|---|
| 43707 | The student must identify the purpose of an author's final paragraphs and provide a detail that supports that purpose. |
| Prompt and Scoring Rubric for item 43707 | Prompt: Explain the purpose of the last two paragraphs in the passage. Then, explain how the author supports that purpose.<br><br>Correct Responses: The author says that it is important to keep inventing.  It is important because zippers can have problems.<br><br><u>2 Point Text:</u><br>Response includes the following correct purpose with a correct support.<br><br>Purpose:<br>    &#9673;  To explain that it is important to keep inventing<br>        &#9673;  To explain that good ideas are only the beginning / coming up with just an idea is easy<br>    &#9673;  To suggest that the reader may invent something faster / easier / cheaper<br>        &#9673;  It could be the reader's new idea that holds things together in the future.<br>    &#9673;  To make the reader think about future inventions<br><br>Support:<br>    &#9673;  Zippers have flaws / problems.<br>        &#9673;  Zippers are imperfect / not perfect.<br>        &#9673;  Zippers break.<br>        &#9673;  People can improve zippers.<br>    &#9673;  Zippers are more complicated than they seem.<br>        &#9673;  Zippers can be tricky.<br>    &#9673;  Inventions are important to the future.<br>    &#9673;  Inventions must work to be useful / it's important for people to come up with inventions that work.<br><br><u>1 Point Text:</u><br>Response includes the correct purpose listed above with an incorrect or missing support.<br><br>Note: The student will NOT receive credit for a correct support without a correct purpose.<br>Sample Answer: The author says it is important to keep inventing.<br><br><u>0 Point Text:</u><br>The response does not meet the criteria required to earn one point. The response indicates inadequate or no understanding of the task and/or the idea or concept needed to answer the item. It may only repeat information given in the test item. The response may provide an incorrect solution/response and the provided supportive information may be irrelevant to the item, or possibly, no other information is shown. The student may have written on a different topic or written, "I don't know." |
| 43964 | The student must write an ending for an unfinished story. Specific details and appropriate word choice/vocabulary are required for full credit. |

| | |
|---|---|
| | |
| Prompt and Scoring Rubric for item 43964 | Prompt:<br>*Read the following paragraphs about a student giving a speech to the rest of her class. Then, complete the task that follows.*<br><br>Sitting and watching Eric give his speech, Kate knew she was next. She dreaded the eventual trek to the front of the room to face her peers and speak. She would have much preferred to read someone else's words, not her own. She felt them so closely, knew them by heart. But what if they disappeared? What if she couldn't remember once she stared into the eyes of those around her?<br><br>"Kate? Kate, it's your turn."<br><br>"Please, can I go to the bathroom?" Kate squeaked, in one final attempt to escape.<br><br>"Once you're finished with your speech. Now come on," the teacher responded.<br><br>Knees shaking, Kate walked to the front of the room. Her stomach turned, twisted, and flipped around and her brain darted from one thought to the next. Her eyes focused on her teacher, then she inhaled and began.<br><br>Now, complete the story.<br><br>Type your answer in the space provided.<br><br>Correct Responses: The first few words were the hardest. Her tongue felt like sandpaper and the words came out slowly, with a rasp and a scratch. But then Kate glanced around at the faces surrounding her. Every gaze was focused on her. But instead of feeling scared, Kate felt empowered. So many people were listening to what she had to say. Kate stood up straighter, spoke louder and felt more free. Her words were important; her fear was not.<br><br>2 Point Text:<br>The response:<br>    • provides appropriate and predominately specific details or evidence<br>    • uses appropriate word choices for intended audience and purpose<br><br>1 Point Text:<br>The response:<br>    • provides mostly general details and evidence, but may include extraneous or loosely related details<br>    • has a limited and predictable vocabulary that may not be consistently appropriate for the intended audience and purpose<br>Sample Answer: Kate's tongue felt like sandpaper and the words came out slowly. Kate stood up straighter, spoke louder and felt more free. Her words were important; her fear was not.<br><br>0 Point Text:<br>The response: |

| | |
|---|---|
| | • includes few supporting details that may be vague, repetitive, incorrect, or interfere with the meaning of the text<br>• has inappropriate vocabulary for the intended audience and purpose |
| 43284 | The student must write an original story using information from the provided sources about games and sports. A full-credit response uses proper style, narrative techniques, and vocabulary. |
| Prompt and Scoring Rubric for item 43284 | Prompt: You have 70 minutes to review your sources, plan, draft, and revise your story. You may refer to the sources. You may also refer to the answer you wrote to the question in Part 1, but you cannot change the answer. Now read your assignment and the information about how your story will be scored; then begin your work.<br><br>Your Assignment<br><br>You have learned about different games people like to play and win from the photo, articles, and stories. You will now write a story from the point of view of someone involved in or watching a game, event, or sport.<br><br>Use the information from the sources about games and sports to write a narrative story. You should present factual information about the activity and also develop characters and a plot.<br><br>Story Scoring<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>Now begin work on your story. Manage your time carefully so that you can:<br><br>• plan your story<br>• write your story<br>• revise and edit your story<br><br>REMEMBER: A well-written narrative story:<br><br>• has a setting, narrative, and/or characters<br>• has a plot with a beginning, middle, and end<br>• uses clear language that suits your purpose<br>• follows rules of writing (spelling, punctuation, and grammar)<br><br>4 Point Text:<br>Establishment of Narrative Focus and Organization<br><br>The narrative, real or imagined, is clearly focused and maintained throughout: |

- effectively establishes a setting, narrator and/or characters, and point of view

The narrative, real or imagined, has an effective plot helping create a sense of unity and completeness:

- consistent use of a variety of transitional strategies to clarify the relationships between and among ideas
- logical sequence of events from beginning to end
- effective opening and closure for audience and purpose

Development/Elaboration

The narrative, real or imagined, provides thorough and effective elaboration using details, dialogue, and description:

- effective use of a variety of narrative techniques that advance the story or illustrate the experience

The narrative, real or imagined, clearly and effectively expresses experiences or events:

- effective use of sensory, concrete, and figurative language clearly advance the purpose

3 Point Text:
Establishment of Narrative Focus and Organization

The narrative, real or imagined, is adequately focused and generally maintained throughout:

- adequately establishes a setting, narrator and/or characters, and/or point of view

The narrative, real or imagined, has an evident plot helping to create a sense of unity and completeness, though there may be minor flaws and some ideas may be loosely connected:

- adequate use of a variety of transitional strategies to clarify the relationships between and among ideas
- adequate sequence of events from beginning to end
- adequate opening and closure for audience and purpose

Development/Elaboration

The narrative, real or imagined, provides adequate elaboration using details, dialogue and description:

- adequate use of a variety of narrative techniques that generally advance the story or illustrate the experience

The narrative, real or imagined, adequately expresses experiences or events:

- adequate use of sensory, concrete, and figurative language generally advance the purpose

2 Point Text:
Establishment of Narrative Focus and Organization

The narrative, real or imagined, is somewhat maintained and may have a minor drift in focus:

- ❿ inconsistently establishes a setting, narrator and/or characters, and/or point of view

The narrative, real or imagined, has an inconsistent plot, and flaws are evident:

- ❿ inconsistent use of transitional strategies and/or little variety
- ❿ uneven sequence of events from beginning to end
- ❿ opening and closure, if present, are weak
- ❿ weak connection among ideas

Development/Elaboration

The narrative, real or imagined, provides uneven, cursory elaboration using partial and uneven details, dialogue, and description:

- ❿ narrative techniques, if present, are uneven and inconsistent

The narrative, real or imagined, unevenly expresses experiences or events:

- ❿ partial or weak use of sensory, concrete, and figurative language that may not advance the purpose

Conventions

The narrative, real or imagined, demonstrates an adequate command of conventions:

- ❿ errors in usage and sentence formation but no systematic pattern of errors is displayed and meaning is not obscured
- ❿ adequate use of punctuation, capitalization, and spelling

1 Point Text:
Establishment of Narrative Focus and Organization

The narrative, real or imagined, may be maintained but may provide little or no focus:

- ❿ may be very brief
- ❿ may have a major drift in focus
- ❿ focus may be confusing or ambiguous

The narrative, real or imagined, has little or no discernible plot:

- ❿ few or no transitional strategies are evident

| | |
|---|---|
| | ✆ frequent extraneous ideas may intrude

Development/Elaboration

The narrative, real or imagined, provides minimal elaboration using little or no details, dialogue, and/or description:

✆ use of narrative techniques is minimal, absent, incorrect, or irrelevant

The narrative, real or imagined, expression of ideas is vague, lacks clarity, or is confusing:

✆ uses limited language
✆ may have little sense of purpose

Conventions

The narrative, real or imagined, demonstrates a partial command of conventions:

✆ frequent errors in usage may obscure meaning
✆ inconsistent use of punctuation, capitalization, and spelling

<u>0 Point Text:</u>
Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing |
| 43334 | The student must use information from the provided documents to write a letter about the importance of visiting the dentist. A full-credit response use proper style, organization, reasoning, and vocabulary. |
| Prompt and Scoring Rubric for item 43334 | Prompt: You have 70 minutes to review your sources, plan, draft, and revise your letter. You may refer to the sources. Read your assignment and the information about how your letter will be scored; then begin your work.

Your Assignment

Your friend tells you he has a dentist's appointment. This is his first dentist appointment and he doesn't know what to expect. You decide to write a letter to your friend informing him of what he can expect at the dentist's office.

In Your Letter

Write a well-organized, multi-paragraph letter explaining why it is important to visit the dentist regularly and what to expect at the dentist's office. Be sure to include details from the articles to support your explanation.

Now begin work on your letter. Manage your time carefully so that you can:
✎ plan your letter
✎ write your letter
✎ revise and edit for a final draft |

Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.

REMEMBER: A well-written informational letter:

- has a clear main idea
- is well-organized and stays on the topic
- provides reasoning and evidence to support your topic
- uses clear language that suits your purpose
- follows rules of writing (spelling, punctuation, and grammar)

4 Point Text:
Statement of Purpose/Focus and Organization
The response is fully sustained, and consistently and purposefully focused:
- controlling idea or main idea of a topic is clearly stated, focused, and strongly maintained
- controlling idea or main idea of a topic is introduced and communicated clearly within the purpose, audience, and task

The response has a clear and effective organizational structure creating a sense of unity and completeness:
- consistent use of a variety of transitional strategies to clarify the relationships between and among ideas
- logical progression of ideas from beginning to end
- effective introduction and conclusion for audience and purpose

Evidence/Elaboration
The response provides thorough and convincing support/evidence for the controlling idea or main idea that includes the effective use of sources, facts, and details:
- use of evidence from sources is integrated, comprehensive, and relevant
- effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:
- use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

3 Point Text:

Statement of Purpose/Focus and Organization
The response is adequately sustained and generally focused:
- controlling idea or main idea of a topic is clear and mostly maintained, though some loosely related material may be present
- some context for the controlling idea or main idea of the topic is adequate within the purpose, audience, and task

The response has an evident organizational structure and a sense of completeness, though there may be minor flaws and some ideas may be loosely connected:
- adequate use of transitional strategies with some variety to clarify the relationships between and

among ideas
- adequate progression of ideas from beginning to end
- adequate introduction and conclusion

Evidence/Elaboration
The response provides adequate support/evidence for controlling idea or main idea that includes the use of sources, facts, and details:
- some evidence from sources is included, though citations may be general or imprecise
- adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language:
- use of domain-specific vocabulary is generally appropriate for the audience and purpose

2 Point Text:
Statement of Purpose/Focus and Organization
The response is somewhat sustained and may have a minor drift in focus:
- may be clearly focused on the controlling or main idea, but is insufficiently sustained, or
- controlling idea or main idea may be unclear and/or somewhat unfocused

The response has an inconsistent organizational structure, and flaws are evident:
- inconsistent use of transitional strategies and/or little variety
- uneven progression of ideas from beginning to end
- conclusion and introduction, if present, are weak

Evidence/Elaboration
The response provides uneven, cursory support/evidence for the controlling idea or main idea that includes partial or uneven use of sources, facts, and details:
- evidence from sources is weakly integrated, and citations, if present, are uneven
- weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:
- use of domain-specific vocabulary that may at times be inappropriate for the audience and purpose

Conventions
The response demonstrates an adequate command of conventions:
- errors in usage and sentence formation may be present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

1 Point Text:

Statement of Purpose/Focus and Organization
The response may be related to the topic but may provide little or no focus:
- may be very brief
- may have a major drift
- focus may be confusing or ambiguous

| | |
|---|---|
| | The response has little or no discernible organizational structure:<br>  • few or no transitional strategies are evident<br>  • frequent extraneous ideas may intrude<br><br>Evidence/Elaboration<br>The response provides minimal support/evidence for the controlling idea or main idea that includes little or no use of sources, facts, and details:<br>  • use of evidence from the source material is minimal, absent, incorrect, or irrelevant<br><br>The response's expression of ideas is vague, lacks clarity, or is confusing:<br>  • uses limited language or domain-specific vocabulary<br>  • may have little sense of audience and purpose<br><br>Conventions<br>The response demonstrates a partial command of conventions:<br>  • errors in usage may obscure meaning<br>  • inconsistent use of punctuation, capitalization, and spelling<br><br>0 Point Text:<br><br>A response gets no credit if it provides no evidence of the ability to structure and write an essay. |
| 43438 | The student must use evidence from the provided documents to write an argumentative essay for or against tourism in national parks. A full-credit response uses proper organization, focus, evidence, style, and vocabulary. |
| Prompt and Scoring Rubric for item 43438 | Prompt: You have 70 minutes to review your sources, plan, draft, and revise your argumentative article. You may refer to the sources. Read your assignment and the information about how your article will be scored; then begin your work.<br><br>Your Assignment<br><br>Your class is planning a field trip to a national park. After researching about the role of the National Park Service, you have been asked by your teacher to write an argumentative article about national parks for the school newspaper.<br><br>Write an article that argues whether the National Park Service should or should not promote tourism for national parks to increase attendance. Be sure that your argument acknowledges both sides of the issue so that people know that you have considered the issue carefully. Support your claim with evidence from the sources. You do not need to use all the sources, only the ones that most effectively support your argument.<br><br>Article Scoring<br><br>Your argumentative article will be scored on the following criteria:<br><br>1. Statement of purpose / focus and organization—How well did you clearly state your claim on the topic and maintain your focus? How well did your ideas logically flow from the introduction to conclusion using |

effective transitions? How well did you stay on topic throughout the article?

2. Elaboration of evidence—How well did you provide evidence from the sources to support your opinions? How well did you elaborate with specific information from the sources you reviewed? How well did you effectively express ideas using precise language that was appropriate for your audience and purpose?

3. Conventions—How well did you follow the rules of usage, punctuation, capitalization, and spelling?

Now begin work on your article. Manage your time carefully so that you can:

- plan your article
- write your article
- revise and edit for a final draft

Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.

REMEMBER: A well-written argumentative article:
- has a clear main idea
- is well-organized and stays on topic
- provides evidence from the sources to support your topic
- uses clear language that suits your purpose
- follows rules of writing (spelling, punctuation, and grammar)

4 Point Text:

Statement of Purpose/Focus and Organization

The response is fully sustained and consistently and purposefully focused:

- claim is clearly stated, focused, and strongly maintained
- alternate or opposing claims are clearly addressed
- claim is introduced and communicated clearly within the purpose, audience, and task

The response has a clear and effective organizational structure creating a sense of unity and completeness:

- consistent use of a variety of transitional strategies to clarify the relationships between and among ideas
- logical progression of ideas from beginning to end
- effective introduction and conclusion for audience and purpose
- strong connections among ideas, with some syntactic variety

Evidence/Elaboration

The response provides thorough and convincing support/evidence for the writer's claim that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:

- use of evidence from sources is integrated, comprehensive, relevant, and concrete
- effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:

- use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

3 Point Text:

Statement of Purpose/Focus and Organization

The response is fully sustained and consistently and purposefully focused:

- claim is clearly stated, focused, and strongly maintained
- alternate or opposing claims are clearly addressed
- claim is introduced and communicated clearly within the purpose, audience, and task

The response has a clear and effective organizational structure creating a sense of unity and completeness:

- consistent use of a variety of transitional strategies to clarify the relationships between and among ideas
- logical progression of ideas from beginning to end
- effective introduction and conclusion for audience and purpose
- strong connections among ideas, with some syntactic variety

Evidence/Elaboration

The response provides thorough and convincing support/evidence for the writer's claim that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:

- use of evidence from sources is integrated, comprehensive, relevant, and concrete
- effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:

- use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

2 Point Text:

Statement of Purpose/Focus and Organization

The response is somewhat sustained and may have a minor drift in focus:

- may be clearly focused on the claim but is insufficiently sustained, or
- claim on the issue may be somewhat unclear and/or unfocused

The response has an inconsistent organizational structure, and flaws are evident:

- inconsistent use of transitional strategies and/or little variety
- uneven progression of ideas from beginning to end
- conclusion and introduction, if present, are weak
- weak connection among ideas

Evidence/Elaboration

The response provides uneven, cursory support/evidence for the writer's claim that includes partial or uneven use of sources, facts, and details. The response achieves little depth:

- evidence from sources is weakly integrated, and citations, if present, are uneven
- weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:

- use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

Conventions

The response demonstrates an adequate command of conventions:

- errors in usage and sentence formation may be present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

<u>1 Point Text:</u>

Statement of Purpose/Focus and Organization

The response may be related to the purpose but may provide little or no focus:

- may be very brief
- may have a major drift
- claim may be confusing or ambiguous

The response has little or no discernible organizational structure:

| | few or no transitional strategies are evident <br> frequent extraneous ideas may intrude |
|---|---|
| | **Evidence/Elaboration** <br><br> The response provides minimal support/evidence for the writer's claim that includes little or no use of sources, facts, and details: <br><br> use of evidence from sources is minimal, absent, incorrect, or irrelevant <br><br> The response's expression of ideas is vague, lacks clarity, or is confusing: <br><br> uses limited language or domain-specific vocabulary <br> may have little sense of audience and purpose <br><br> **Conventions** <br><br> The response demonstrates a partial command of conventions: <br><br> errors in usage may obscure meaning <br> inconsistent use of punctuation, capitalization, and spelling <br><br> <u>0 Point Text:</u> <br><br> Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing |
| 43469 | The student must write an orginal story about traveling to the Moon, using information from the provided documents. A full-credit responses uses proper narrative techniques, details, style, and vocabulary. |
| Prompt and Scoring Rubric for item 43469 | Prompt: You will now have 70 minutes to review your documents, plan, draft, and revise you narrative story. You may refer to the documents. You may also refer to the answer you wrote to the question in Part 1, but you cannot change that answer. Now read your assignment and the information about how your story will be scored; then begin your work. <br><br> Your Assignment <br><br> Your class is studying a unit about science fiction literature. You have been learning about how science fiction stories are based on new understandings in science. <br><br> Your assignment is to write a short science fiction story about traveling to the Moon or living on the Moon using current scientific knowledge to shape the story. You should include details from the source material for your narrative story. You do not need to use all the documents, only the ones that best support the details in your narrative story. <br><br> Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen. <br><br> Now begin work on your story. Manage your time carefully so that you can: |

- plan your story
- write your story
- revise and edit your story

Narrative Scoring

Your narrative story will be scored on the following criteria:

- develops a setting and characters
- uses a point of view
- has a plot with a beginning, middle, and end
- uses details and dialogue
- uses clear language that suits your purpose
- follows rules of writing (spelling, punctuation, and grammar)

4 Point Text:

Establishment of Narrative Focus and Organization
The narrative, real or imagined, is clearly focused and maintained throughout:
- effectively establishes a setting, narrator and/or characters, and point of view

The narrative, real or imagined, has an effective plot helping create unity and completeness:
- effective, consistent use of a variety of transitional strategies
- logical sequence of events from beginning to end
- effective opening and closure for audience and purpose

Development/Elaboration
The narrative, real or imagined, provides thorough and effective elaboration using details, dialogue, and description:
- effective use of a variety of narrative techniques that advance the story or illustrate the experience

The narrative, real or imagined, clearly and effectively expresses experiences or events:
- effective use of sensory, concrete, and figurative language clearly advance the purpose

3 Point Text:

Establishment of Narrative Focus and Organization
The narrative, real or imagined, is adequately focused and generally maintained throughout:
- adequately establishes a setting, narrator and/or characters, and point of view

The narrative, real or imagined, has an evident plot helping create a sense of unity and completeness, though there may be minor flaws and some ideas may be loosely connected:
- adequate use of a variety of transitional strategies
- adequate sequence of events from beginning to end
- adequate opening and closure for audience and purpose

Development/Elaboration
The narrative, real or imagined, provides adequate elaboration using details, dialogue, and description:
- adequate use of a variety of narrative techniques that generally advance the story or illustrate the experience

The narrative, real or imagined, adequately expresses experiences or events:
- adequate use of sensory, concrete, and figurative language generally advance the purpose

2 Point Text:

Establishment of Narrative Focus and Organization
The narrative, real or imagined, is somewhat maintained and may have a minor drift in focus:
- inconsistently establishes a setting, narrator and/or characters, and point of view

The narrative, real or imagined, has an inconsistent plot, and flaws are evident:
- inconsistent use of basic transitional strategies with little variety
- uneven sequence of events from beginning to end
- opening and closure, if present, are weak
- weak connection among ideas

Development/Elaboration
The narrative, real or imagined, provides uneven, cursory elaboration using partial and uneven details, dialogue, and description:
- narrative techniques, if present, are uneven and inconsistent

The narrative, real or imagined, unevenly expresses experiences or events:
- partial or weak use of sensory, concrete, and figurative language that may not advance the purpose

The narrative, real or imagined, demonstrates an adequate command of conventions:
- errors in usage and sentence formation, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

1 Point Text:

Establishment of Narrative Focus and Organization
The narrative, real or imagined, may be maintained but may provide little or no focus:
- may be very brief
- may have a major drift
- focus may be confusing or ambiguous

The narrative, real or imagined, has little or no discernible plot:
- few or no transitional strategies are evident
- frequent extraneous ideas may intrude

| | |
|---|---|
| | Development/Elaboration<br>The narrative, real or imagined, provides minimal elaboration using little or no details, dialogue, and description:<br>    ❿  use of narrative techniques is minimal, absent, in error, or irrelevant<br><br>The narrative, real or imagined, expression of ideas is vague, lacks clarity, or is confusing:<br>    ❿  uses limited language<br>    ❿  may have little sense of purpose<br><br>Conventions<br>The narrative, real or imagined, demonstrates a partial command of conventions:<br>    ❿  frequent errors in usage may obscure meaning<br>    ❿  inconsistent use of punctuation, capitalization, and spelling<br><br>0 Point Text:<br><br>Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing |
| 43479 | The student must use evidence from the provided documents to write an argumentative essay for attending college either in a small town or a big city. A full-credit response uses proper organization, focus, evidence, style, and vocabulary. |
| Prompt and Scoring Rubric for item 43479 | Prompt: The editor of your school newspaper has asked you to write an argumentative article about whether students should go to college in a small town or a big city. After you analyze all of the sources, determine which view you support. Compose a full-length argumentative essay in support of your view. In your writing, use logical reasoning and evidence from the sources to support your claim and to refute counterarguments.<br><br>Now begin work on your argumentative article. Manage your time carefully so that you can:<br><br>    ✎  plan your article<br>    ✎  write your article<br>    ✎  revise and edit for a final draft<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>REMEMBER: A well-written article:<br><br>    ✎  has a clear main idea<br>    ✎  is well-organized and stays on the topic<br>    ✎  provides evidence from the sources to support your topic<br>    ✎  uses clear language that suits your purpose<br>    ✎  follows rules of writing (spelling, punctuation, and grammar)<br><br>4 Point Text:<br>The response is fully sustained and consistently and purposefully focused: |

- ⦾ claim is clearly stated, focused and strongly maintained
- ⦾ alternate or opposing claims are clearly addressed
- ⦾ claim is introduced and communicated clearly within the context

The response has a clear and effective organizational structure creating unity and completeness:
- ⦾ effective, consistent use of a variety of transitional strategies
- ⦾ logical progression of ideas from beginning to end
- ⦾ effective introduction and conclusion for audience and purpose
- ⦾ strong connections among ideas, with some syntactic variety

The response provides thorough and convincing support/evidence for the writer's claim that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:
- ⦾ use of evidence from sources is smoothly integrated, comprehensive, relevant, and concrete
- ⦾ effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:
- ⦾ use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

The response demonstrates a strong command of conventions:
- ⦾ few, if any, errors are present in usage and sentence formation
- ⦾ effective and consistent use of punctuation, capitalization, and spelling

3 Point Text:
The response is adequately sustained and generally focused:
- ⦾ claim is clear and for the most part maintained, though some loosely related material may be present
- ⦾ context provided for the claim is adequate

The response has an evident organizational structure and a sense of completeness, though there may be minor flaws and some ideas may be loosely connected:
- ⦾ adequate use of transitional strategies with some variety
- ⦾ adequate progression of ideas from beginning to end
- ⦾ adequate introduction and conclusion
- ⦾ adequate, if slightly inconsistent, connection among ideas

The response provides adequate support/evidence for writer's claim that includes the use of sources, facts, and details. The response achieves some depth and specificity but is predominantly general:
- ⦾ some evidence from sources is integrated, though citations may be general or imprecise
- ⦾ adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language:
- ⦾ use of domain-specific vocabulary is generally appropriate for the audience and purpose

The response demonstrates an adequate command of conventions:
- ⦾ some errors in usage and sentence formation may be present, but no systematic pattern of

errors is displayed
- adequate use of punctuation, capitalization, and spelling

2 Point Text:
The response is somewhat sustained and may have a minor drift in focus:
- may be clearly focused on the claim but is insufficiently sustained
- claim on the issue may be somewhat unclear and unfocused

The response has an inconsistent organizational structure, and flaws are evident:
- inconsistent use of basic transitional strategies with little variety
- uneven progression of ideas from beginning to end
- conclusion and introduction, if present, are weak
- weak connection among ideas

The response provides uneven, cursory support/evidence for the writer's claim that includes partial or uneven use of sources, facts, and details, and achieves little depth:
- evidence from sources is weakly integrated, and citations, if present, are uneven
- weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:
- use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

The response demonstrates a partial command of conventions:
- frequent errors in usage may obscure meaning
- inconsistent use of punctuation, capitalization, and spelling

1 Point Text:
The response may be related to the purpose but may offer little relevant detail:
- may be very brief
- may have a major drift
- claim may be confusing or ambiguous

The response has little or no discernible organizational structure:
- few or no transitional strategies are evident
- frequent extraneous ideas may intrude

The response provides minimal support/evidence for the writer's claim that includes little or no use of sources, facts, and details:
- use of evidence from sources is minimal, absent, in error, or irrelevant

The response expression of ideas is vague, lacks clarity, or is confusing:
- uses limited language or domain-specific vocabulary
- may have little sense of audience and purpose

The response demonstrates a lack of command of conventions:
- errors are frequent and severe and meaning is often obscure

| | |
|---|---|
| | **0 Point Text:**<br>A response gets no credit if it provides no evidence of the ability to establish and support a formal argumentative claim. |
| 43504 | The student must write an opinion essay about the region of the US they would like to live in. A full-credit response includes details from the provided documents and uses proper organization, style, and vocabulary. |
| Prompt and Scoring Rubric for item 43504 | Prompt: You have 70 minutes to review your sources, plan, draft, and revise your essay. Read your assignment and the information about how your essay will be scored; then begin your work.<br><br>Your Assignment<br><br>You have read three documents about different regions in the United States. You have been asked to give an opinion about these regions.<br><br>Write an opinion essay about which region you would want to live in. State your opinion in the essay, using reasons and supporting details from the sources to explain what you like about this region and why.<br><br>In Your Essay<br><br>Write a well-organized essay that develops your opinion about which region you would want to live in. Be sure to include reasons and details from the sources to support your opinion.<br><br>Now begin work on your opinion essay. Manage your time carefully so that you can:<br><br>   &#x25b7; plan your essay<br>   &#x25b7; write your essay<br>   &#x25b7; revise and edit for a final draft<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>REMEMBER: A well-written opinion essay:<br><br>   &#x25b7; has a clear opinion<br>   &#x25b7; is well-organized and stays on the topic<br>   &#x25b7; provides evidence and details from the sources to support your opinion<br>   &#x25b7; uses clear language that suits your purpose<br>   &#x25b7; follows rules of writing (spelling, punctuation, and grammar)<br><br><br>**4 Point Text:**<br>Statement of Purpose/Focus and Organization<br>The response is fully sustained and consistently and purposefully focused:<br>   &#x24ea; opinion is clearly stated, focused, and strongly maintained<br>   &#x24ea; opinion is communicated clearly within the purpose, audience, and task |

The response has a clear and effective organizational structure creating a sense of unity and completeness:
- consistent use of a variety of transitional strategies to clarify the relationships between and among ideas
- logical progression of ideas from beginning to end
- effective introduction and conclusion for audience and purpose

Evidence/Elaboration

The response provides thorough and convincing support/evidence for the writer's opinion that includes the effective use of sources, facts, and details:
- use of evidence from sources is integrated, comprehensive, and relevant
- effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:
- use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

3 Point Text:

Statement of Purpose/Focus and Organization

The response is adequately sustained and generally focused:
- opinion is clear and mostly maintained, though some loosely related material may be present
- context provided for the opinion is adequate within the purpose, audience, and task

The response has a recognizable organizational structure, and a sense of completeness, though there may be minor flaws and some ideas may be loosely connected:
- adequate use of transitional strategies with some variety to clarify the relationships between and among ideas
- adequate progression of ideas from beginning to end
- adequate introduction and conclusion

Evidence/Elaboration

The response provides adequate support/evidence for the writer's opinion that includes the use of sources, facts, and details:
- some evidence from sources is included, though citations may be general or imprecise
- adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language:
- use of domain-specific vocabulary is generally appropriate for the audience and purpose

2 Point Text:

Statement of Purpose/Focus and Organization

The response is somewhat sustained and may have a minor drift in focus:
- may be clearly focused on the opinion but is insufficiently sustained, or
- opinion on the issue may be somewhat unclear and/or unfocused

The response has an inconsistent organizational structure, and flaws are evident:
- inconsistent use of transitional strategies and/or little variety
- uneven progression of ideas from beginning to end
- conclusion and introduction, if present, are weak

Evidence/Elaboration

The response provides uneven, cursory support/evidence for the writer's opinion that includes partial or uneven use of sources, facts, and details:
- evidence from sources is weakly integrated, and citations, if present, are uneven
- weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:
- use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

Conventions

The response demonstrates an adequate command of conventions:
- errors in usage and sentence formation are present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

<u>1 Point Text:</u>

Statement of Purpose/Focus and Organization

The response may be related to the opinion but may provide little or no focus:
- may be very brief
- may have a major drift
- opinion may be confusing or ambiguous

The response has little or no discernible organizational structure:
- few or no transitional strategies are evident
- frequent extraneous ideas may intrude

Evidence/Elaboration

The response provides minimal support/evidence for the writer's opinion that includes little or no use of sources, facts, and details:
- use of evidence from sources is minimal, absent, incorrect, or irrelevant

The response's expression of ideas is vague, lacks clarity, or is confusing:
- uses limited language or domain-specific vocabulary
- may have little sense of audience and purpose

Conventions

The response demonstrates a partial command of conventions:
- errors in usage may obscure meaning
- inconsistent use of punctuation, capitalization, and spelling

| | |
|---|---|
| | <u>0 Point Text:</u><br><br>Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing. |
| 43632 | Using information from the provided documents, the student must write a report about different methods of studying. A full-credit response uses proper organization, evidence, style, and vocabulary. |
| Prompt and Scoring Rubric for item 43632 | Prompt: Imagine you are the student who wrote the journal entry. You research better study skills and decide to share them with your friend Megan. Your guidance counselor hears about what you have learned and asks if you could write an informational report for all the students in your school. You accept her request.<br><br>Write an informational report about different methods of improving memory and learning. Explain how effective these methods are and why. Then, recommend a general plan for all students. Support your report with evidence from the sources you have examined.<br><br>Now begin work on your informational report. Manage your time carefully so that you can:<br><br>    ✍   plan your report<br>    ✍   write your report<br>    ✍   revise and edit for a final draft<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>REMEMBER: A well-written report:<br><br>    ✍   has a clear main idea<br>    ✍   is well-organized and stays on the topic<br>    ✍   provides evidence from the sources to support your topic<br>    ✍   uses clear language that suits your purpose<br>    ✍   follows rules of writing (spelling, punctuation, and grammar)<br><br><u>4 Point Text:</u><br><br>The response is fully sustained, and consistently and purposefully focused:<br>    ✍   controlling idea or main idea of a topic is clearly stated, focused, and strongly maintained<br>    ✍   controlling idea or main idea of a topic is introduced and communicated clearly within the purpose, audience, and task<br><br>The response has a clear and effective organizational structure creating a sense of unity and completeness:<br>    ✍   consistent use of a variety of transitional strategies to clarify the relationships between and among ideas<br>    ✍   logical progression of ideas from beginning to end<br>    ✍   effective introduction and conclusion for audience and purpose<br>    ✍   strong connections among ideas, with some syntactic variety |

The response provides thorough and convincing support/evidence for the controlling idea or main idea that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:

- use of evidence from sources is integrated, comprehensive, relevant, and concrete
- effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:

- use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

The response demonstrates an adequate command of conventions:

- errors in usage and sentence formation are present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

3 Point Text:

The response is adequately sustained and generally focused:

- controlling idea or main idea of a topic is clear mostly maintained, though some loosely related material may be present
- some context for the controlling idea or main idea of the topic is adequate within the purpose, audience, and task

The response has an evident organizational structure and a sense of completeness, though there may be minor flaws and some ideas may be loosely connected:

- adequate use of transitional strategies with some variety between and among ideas
- adequate progression of ideas from beginning to end
- adequate introduction and conclusion
- adequate, if slightly inconsistent, connection among ideas

The response provides adequate support/evidence for the controlling idea or main idea that includes the use of sources, facts, and details:

- some evidence from sources is included, though citations may be general or imprecise
- adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language :

- use of domain-specific vocabulary is generally appropriate for the audience and purpose

The response demonstrates an adequate command of conventions:

- errors in usage and sentence formation are present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

2 Point Text:

The response is somewhat sustained and may have a minor drift in focus:

- may be clearly focused on the controlling or main idea but is insufficiently sustained, or

&#x25A1; controlling idea or main idea may be unclear and/or somewhat unfocused

The response has an inconsistent organizational structure, and flaws are evident:
- &#x25A1; inconsistent use of transitional strategies and/or little variety
- &#x25A1; uneven progression of ideas from beginning to end
- &#x25A1; conclusion and introduction, if present, are weak
- &#x25A1; weak connection among ideas

The response provides uneven, cursory support/evidence for the controlling idea or main idea that includes partial or uneven use of sources, facts, and details. The response achieves little depth:
- &#x25A1; evidence from sources is weakly integrated, and citations, if present, are uneven

The response expresses ideas unevenly, using simplistic language:
- &#x25A1; use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

The response demonstrates a lack of command of conventions:
- &#x25A1; errors are frequent and severe and meaning is often obscure

<u>1 Point Text:</u>

The response may be related to the topic but may provide little or no focus:
- &#x25A1; may be very brief
- &#x25A1; may have a major drift
- &#x25A1; focus may be confusing or ambiguous

The response has little or no discernible organizational structure:
- &#x25A1; few or no transitional strategies are evident
- &#x25A1; frequent extraneous ideas may intrude

The response provides minimal support/evidence for the controlling idea or main idea that includes little or no use of sources, facts, and details:
- &#x25A1; use of evidence from sources is minimal, absent, incorrect, or irrelevant

The response expression of ideas is vague, lacks clarity, or is confusing:
- &#x25A1; uses limited language or domain-specific vocabulary
- &#x25A1; may have little sense of audience and purpose

The response demonstrates a lack of command of conventions:
- &#x25A1; errors are frequent and severe and meaning is often obscure

<u>0 Point Text:</u>

Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing

| 43635 | The student must write an editorial about the pros and cons of cell phones in daily life. A full-credit response includes information from the provided documents and uses proper organization, style, and |
|---|---|

| | vocabulary. |
|---|---|
| Prompt and Scoring Rubric for item 43635 | Prompt: You are interested in pursuing a career in journalism and decide to apply for a position with your school newspaper. The editor-in-chief asks you to submit a viewpoint editorial for consideration. You decide to write about cell phones.<br><br>Write an argumentative essay that evaluates the pros and cons of cell phone use and states whether cell phones make daily life better or worse. Make sure to address potential counterarguments in your essay and support your claim with the sources you have examined.<br><br>Now begin work on your argumentative essay. Manage your time carefully so that you can:<br><br>  &#x2767; plan your essay<br>  &#x2767; write your essay<br>  &#x2767; revise and edit for a final draft<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>REMEMBER: A well-written argumentative essay:<br><br>  &#x2767; has a clear main idea<br>  &#x2767; is well-organized and stays on the topic<br>  &#x2767; provides evidence from the sources to support your topic<br>  &#x2767; uses clear language that suits your purpose<br>  &#x2767; follows rules of writing (spelling, punctuation, and grammar)<br><br>4 Point Text:<br><br>The response provides thorough and convincing support/evidence for the writer's claim that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:<br>  &#x2767; use of evidence from sources is integrated, comprehensive, relevant, and concrete<br>  &#x2767; effective use of a variety of elaborative techniques<br><br>The response clearly and effectively expresses ideas, using precise language:<br>  &#x2767; use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose<br><br>The response demonstrates an adequate command of conventions:<br>  &#x2767; errors in usage and sentence formation may be present, but no systematic pattern of errors is displayed and meaning is not obscured<br>  &#x2767; adequate use of punctuation, capitalization, and spelling<br><br>3 Point Text:<br><br>The response provides adequate support/evidence for the writer's claim that includes the use of sources, facts, and details.  The response achieves some depth and specificity but is predominantly |

general:
- some evidence from sources is included, though citations may be general or imprecise
- adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language:
- use of domain-specific vocabulary is generally appropriate for the audience and purpose

The response demonstrates an adequate command of conventions:
- errors in usage and sentence formation may be present, but no systematic pattern of errors is displayed and meaning is not obscured
- adequate use of punctuation, capitalization, and spelling

2 Point Text:

The response provides uneven, cursory support/evidence for the writer's claim that includes partial or uneven use of sources, facts, and details. The response achieves little depth:
- evidence from sources is weakly integrated, and citations, if present, are uneven
- weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:
- use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

The response demonstrates a partial command of conventions:
- errors in usage may obscure meaning
- inconsistent use of punctuation, capitalization, and spelling

1 Point Text:

The response provides minimal support/evidence for the writer's claim that includes little or no use of sources, facts, and details:
- use of evidence from sources is minimal, absent, incorrect, or irrelevant

The response expression of ideas is vague, lacks clarity, or is confusing:
- uses limited language or domain-specific vocabulary
- may have little sense of audience and purpose

The response demonstrates a partial command of conventions:
- errors in usage may obscure meaning
- inconsistent use of punctuation, capitalization, and spelling

0 Point Text:

A response gets no credit if it provides no evidence of the ability to structure and write an essay.

| 43703 | The student must use information in the provided documents to write an essay explaining what constellations can tell you about different cultures. A full-credit response uses proper organization, style, and vocabulary. |
|---|---|

| | |
|---|---|
| Prompt and Scoring Rubric for item 43703 | Prompt: You have 70 minutes to review your sources, plan, draft, and revise your essay. You may refer to the sources. Read your assignment and the information about how your essay will be scored; then begin your work.<br><br>Your Assignment<br><br>Your class is studying a unit on the stars. Your teacher has asked you to write an essay based on your research.<br><br>Write an essay explaining what the names of the constellations can tell you about the people who named them and the attitudes people have had about the stars through the years. Include details from the sources in your essay. You do not need to use all the sources, only the ones that most effectively support the main ideas in your explanatory essay.<br><br>Essay Scoring<br><br>Your explanatory essay will be scored on the following criteria:<br><br>1. Statement of purpose / focus and organization—How well did you clearly state your main idea? How well did your ideas logically flow from the introduction to conclusion using effective transitions? How well did you stay on topic throughout the essay?<br>2. Elaboration of evidence—How well did you provide evidence from the sources to support your main ideas? How well did you elaborate with specific information from the sources you reviewed? How well did you effectively express ideas using precise language that was appropriate for your audience and purpose?<br>3. Conventions—How well did you follow the rules of usage, punctuation, capitalization, and spelling?<br><br>Now begin work on your essay. Manage your time carefully so that you can:<br>   ❿  plan your essay<br>   ❿  write your essay<br>   ❿  revise and edit for a final draft<br><br>Type your response in the space provided. Write as much as you need to fulfill the requirements of the task; you are not limited by the size of the response area on the screen.<br><br>4 Point Text:<br>Statement of Purpose/Focus and Organization<br>The response is fully sustained, and consistently and purposefully focused:<br>   ❿  controlling idea or main idea of a topic is clearly stated, focused, and strongly maintained<br>   ❿  controlling idea or main idea of a topic is introduced and communicated clearly within the purpose, audience, and task<br><br>The response has a clear and effective organizational structure creating a sense of unity and completeness:<br>   ❿  consistent use of a variety of transitional strategies to clarify the relationships between and among ideas.<br>   ❿  logical progression of ideas from beginning to end |

- ◉ effective introduction and conclusion for audience and purpose
- ◉ strong connections among ideas, with some syntactic variety

Evidence/Elaboration
The response provides thorough and convincing support/evidence for the controlling idea or main idea that includes the effective use of sources, facts, and details. The response achieves substantial depth that is specific and relevant:
- ◉ use of evidence from sources is integrated, comprehensive, relevant, and concrete
- ◉ effective use of a variety of elaborative techniques

The response clearly and effectively expresses ideas, using precise language:
- ◉ use of academic and domain-specific vocabulary is clearly appropriate for the audience and purpose

3 Point Text:
Statement of Purpose/Focus and Organization
The response is adequately sustained and generally focused:
- ◉ controlling idea or main idea of a topic is clear mostly maintained, though some loosely related material may be present
- ◉ some context for the controlling idea or main idea of the topic is adequate within the purpose, audience, and task

The response has an evident organizational structure and a sense of completeness, though there may be minor flaws and some ideas may be loosely connected:
- ◉ adequate use of transitional strategies with some variety between and among ideas
- ◉ adequate progression of ideas from beginning to end
- ◉ adequate introduction and conclusion
- ◉ adequate, if slightly inconsistent, connection among ideas

Evidence/Elaboration
The response provides adequate support/evidence for the controlling idea or main idea that includes the use of sources, facts, and details:
- ◉ some evidence from sources is included, though citations may be general or imprecise
- ◉ adequate use of some elaborative techniques

The response adequately expresses ideas, employing a mix of precise with more general language:
- ◉ use of domain-specific vocabulary is generally appropriate for the audience and purpose

2 Point Text:
Statement of Purpose/Focus and Organization
The response is somewhat sustained and may have a minor drift in focus:
- ◉ may be clearly focused on the controlling or main idea but is insufficiently sustained, or
- ◉ controlling idea or main idea may be unclear and/or somewhat unfocused

The response has an inconsistent organizational structure, and flaws are evident:
- ◉ inconsistent use of transitional strategies and/or little variety
- ◉ uneven progression of ideas from beginning to end

- ❿ conclusion and introduction, if present, are weak
- ❿ weak connection among ideas

Evidence/Elaboration
The response provides uneven, cursory support/evidence for the controlling idea or main idea that includes partial or uneven use of sources, facts, and details. The response achieves little depth:
- ❿ evidence from sources is weakly integrated, and citations, if present, are uneven
- ❿ weak or uneven use of elaborative techniques

The response expresses ideas unevenly, using simplistic language:
- ❿ use of domain-specific vocabulary may at times be inappropriate for the audience and purpose

Conventions
The response demonstrates an adequate command of conventions:
- ❿ errors in usage and sentence formation are present, but no systematic pattern of errors is displayed and meaning is not obscured
- ❿ adequate use of punctuation, capitalization, and spelling

1 Point Text:
Statement of Purpose/Focus and Organization
The response may be related to the topic but may provide little or no focus:
- ❿ may be very brief
- ❿ may have a major drift
- ❿ focus may be confusing or ambiguous

The response has little or no discernible organizational structure:
- ❿ few or no transitional strategies are evident
- ❿ frequent extraneous ideas may intrude

Evidence/Elaboration
The response provides minimal support/evidence for the controlling idea or main idea that includes little or no use of sources, facts, and details:
- ❿ use of evidence from sources is minimal, absent, incorrect, or irrelevant

The response's expression of ideas is vague, lacks clarity, or is confusing:
- ❿ uses limited language or domain-specific vocabulary
- ❿ may have little sense of audience and purpose

Conventions
The response demonstrates a lack of command of conventions:
- ❿ errors are frequent and severe and meaning is often obscure

0 Point Text:

Insufficient, illegible, in a language other than English, incoherent, off-topic, or off-purpose writing

Item Analysis Results by Grade and Form

## Grade 4 Mathematics

Table A1. Grade 4 Mathematics: Item Analysis Summary Statistics

| N=1062 | Item Difficulty | Item Discrimination |
|--------|-----------------|---------------------|
| Mean | 0.42 | 0.55 |
| SD | 0.30 | 0.16 |
| Min | 0.02 | 0.15 |
| Max | 0.88 | 0.76 |

Table A2. Grade 4 Mathematics: Distribution of DIF Item Categorizations for Selected Groups

| | LEP vs. non-LEP | Female vs. Male | Hispanic vs. White |
|---|---|---|---|
| | Number of Students | | |
| Focal | 151 | 537 | 234 |
| Reference | 911 | 525 | 455 |
| | Number of Items | | |
| C+ | 2 | 0 | 1 |
| B+ | 2 | 1 | 1 |
| A+ | 5 | 10 | 9 |
| A– | 9 | 8 | 5 |
| B– | 2 | 1 | 4 |
| C– | 0 | 0 | 0 |

## Grade 7 Mathematics

Table B1. Grade 7 Mathematics: Item Analysis Summary Statistics

| N= 710 | Item Difficulty | Item Discrimination |
|--------|-----------------|---------------------|
| Mean | 0.29 | 0.55 |
| SD | 0.14 | 0.17 |
| Min | 0.03 | 0.19 |
| Max | 0.53 | 0.91 |

Table B2. Grade 7 Mathematics: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|--------|-----------------|--------------------|
| | Number of Students | |
| Focal | 367 | 117 |
| Reference | 343 | 342 |
| | Number of Items | |
| C+ | 0 | 1 |
| B+ | 2 | 1 |
| A+ | 7 | 7 |
| A− | 11 | 8 |
| B− | 0 | 3 |
| C− | 0 | 0 |

## Grade 11 Mathematics

Table C1. Grade 11 Mathematics: Item Analysis Summary Statistics

| N=570 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.25 | 0.59 |
| SD | 0.18 | 0.25 |
| Min | 0.00 | -0.03 |
| Max | 0.69 | 1.00 |

Table C2. Grade 11 Mathematics: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 299 | 107 |
| Reference | 271 | 316 |
| | Number of Items | |
| C+ | 0 | 0 |
| B+ | 0 | 0 |
| A+ | 10 | 11 |
| A– | 8 | 9 |
| B– | 2 | 0 |
| C– | 0 | 0 |

## Grade 4 Reading

Table D1. Grade 4 Reading: Item Analysis Summary Statistics

| N=1057 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.45 | 0.51 |
| SD | 0.16 | 0.14 |
| Min | 0.21 | 0.15 |
| Max | 0.82 | 0.68 |

Table D2. Grade 4 Reading: Distribution of DIF Item Categorizations for Selected Groups

| | LEP vs. non-LEP | Female vs. Male | Hispanic vs. White |
|---|---|---|---|
| | Number of Students | | |
| Focal | 150 | 515 | 232 |
| Reference | 907 | 542 | 474 |
| | Number of Items | | |
| C+ | 0 | 0 | 0 |
| B+ | 0 | 2 | 1 |
| A+ | 8 | 7 | 5 |
| A– | 8 | 7 | 12 |
| B– | 2 | 2 | 0 |
| C– | 0 | 0 | 0 |

**Grade 7 Reading**

Table E1. Grade 7 Reading: Item Analysis Summary Statistics

| N=706 | Item Difficulty | Item Discrimination |
|-------|-----------------|---------------------|
| Mean | 0.46 | 0.49 |
| SD | 0.20 | 0.17 |
| Min | 0.19 | 0.06 |
| Max | 0.75 | 0.70 |

Table E2. Grade 7 Reading: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 330 | 128 |
| Reference | 376 | 335 |
| | Number of Items | |
| C+ | 0 | 1 |
| B+ | 4 | 0 |
| A+ | 7 | 9 |
| A– | 2 | 6 |
| B– | 4 | 1 |
| C– | 1 | 1 |

## Grade 11 Reading

Table F1. Grade 11 Reading: Item Analysis Summary Statistics

| N=580 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.41 | 0.45 |
| SD | 0.20 | 0.22 |
| Min | 0.06 | -0.30 |
| Max | 0.70 | 0.65 |

Table F2. Grade 11 Reading: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 303 | 118 |
| Reference | 277 | 320 |
| | Number of Items | |
| C+ | 0 | 0 |
| B+ | 0 | 0 |
| A+ | 10 | 7 |
| A− | 7 | 11 |
| B− | 1 | 0 |
| C− | 0 | 0 |

**Grade 4 Writing Form A**

Table G1. Grade 4 Writing: Item Analysis Summary Statistics

| N=1047 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.43 | 0.50 |
| SD | 0.16 | 0.19 |
| Min | 0.17 | 0.14 |
| Max | 0.82 | 0.81 |

Table G2. Grade 4 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | LEP vs. non-LEP | Female vs. Male | Hispanic vs. White |
|---|---|---|---|
| | Number of Students | | |
| Focal | 124 | 497 | 237 |
| Reference | 923 | 550 | 481 |
| | Number of Items | | |
| C+ | 0 | 0 | 0 |
| B+ | 4 | 0 | 1 |
| A+ | 5 | 9 | 4 |
| A– | 7 | 6 | 12 |
| B– | 2 | 3 | 1 |
| C– | 0 | 0 | 0 |

**Grade 4 Writing Form B**

Table H1. Grade 4 Writing: Item Analysis Summary Statistics

| N=1061 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.44 | 0.49 |
| SD | 0.16 | 0.18 |
| Min | 0.18 | 0.18 |
| Max | 0.82 | 0.82 |

Table H2. Grade 4 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | LEP vs. non-LEP | Female vs. Male | Hispanic vs. White |
|---|---|---|---|
| | Number of Students | | |
| Focal | 145 | 556 | 243 |
| Reference | 916 | 505 | 491 |
| | Number of Items | | |
| C+ | 0 | 0 | 0 |
| B+ | 0 | 2 | 2 |
| A+ | 9 | 5 | 6 |
| A− | 9 | 8 | 9 |
| B− | 1 | 4 | 2 |
| C− | 0 | 0 | 0 |

**Grade 4 Writing Form C**

Table I1. Grade 4 writing: Item Analysis Summary Statistics

| N=1042 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.41 | 0.48 |
| SD | 0.18 | 0.20 |
| Min | 0.15 | 0.08 |
| Max | 0.85 | 0.84 |

Table I2. Grade 4 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | LEP vs. non-LEP | Female vs. Male | Hispanic vs. White |
|---|---|---|---|
| | Number of Students | | |
| Focal | 146 | 547 | 229 |
| Reference | 896 | 495 | 488 |
| | Number of Items | | |
| C+ | 0 | 0 | 0 |
| B+ | 0 | 3 | 2 |
| A+ | 9 | 6 | 8 |
| A– | 8 | 7 | 7 |
| B– | 2 | 3 | 1 |
| C– | 0 | 0 | 1 |

**Grade 7 Writing Form A**

Table J1. Grade 7 Writing: Item Analysis Summary Statistics

| N=698 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.48 | 0.48 |
| SD | 0.22 | 0.17 |
| Min | 0.19 | 0.21 |
| Max | 0.89 | 0.81 |

Table J2. Grade 7 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 339 | 118 |
| Reference | 358 | 341 |
| | Number of Items | |
| C+ | 0 | 0 |
| B+ | 3 | 0 |
| A+ | 4 | 8 |
| A− | 11 | 7 |
| B− | 1 | 2 |
| C− | 0 | 2 |

**Grade 7 Writing Form B**

Table K1. Grade 7 Writing: Item Analysis Summary Statistics

| N=703 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.50 | 0.51 |
| SD | 0.21 | 0.18 |
| Min | 0.21 | 0.16 |
| Max | 0.89 | 0.79 |

Table K2. Grade 7 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | *Number of Students* | |
| Focal | 361 | 123 |
| Reference | 342 | 349 |
| | *Number of Items* | |
| C+ | 1 | 0 |
| B+ | 0 | 3 |
| A+ | 8 | 3 |
| A− | 9 | 10 |
| B− | 1 | 2 |
| C− | 0 | 1 |

**Grade 7 Writing Form C**

Table L1. Grade 7 Writing: Item Analysis Summary Statistics

| N=695 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.49 | 0.49 |
| SD | 0.21 | 0.17 |
| Min | 0.20 | 0.22 |
| Max | 0.89 | 0.81 |

Table L2. Grade 7 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male |
|---|---|
| | Number of Students |
| Focal | 356 |
| Reference | 339 |
| | Number of Items |
| C+ | 0 |
| B+ | 0 |
| A+ | 9 |
| A− | 7 |
| B− | 3 |
| C− | 0 |

## Grade 11 Writing Form A

Table M1. Grade 11 Writing: Item Analysis Summary Statistics

| N=573 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.52 | 0.47 |
| SD | 0.17 | 0.23 |
| Min | 0.06 | -0.27 |
| Max | 0.79 | 0.79 |

Table M2. Grade 11 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 309 | 111 |
| Reference | 264 | 313 |
| | Number of Items | |
| C+ | 0 | 1 |
| B+ | 1 | 1 |
| A+ | 12 | 8 |
| A– | 3 | 8 |
| B– | 2 | 0 |
| C– | 1 | 1 |

**Grade 11 Writing Form B**

Table N1. Grade 11 Writing: Item Analysis Summary Statistics

| N=573 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.52 | 0.46 |
| SD | 0.17 | 0.25 |
| Min | 0.05 | -0.21 |
| Max | 0.80 | 0.80 |

Table N2. Grade 11 Writing: Distribution of DIF item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 285 | 125 |
| Reference | 288 | 309 |
| | Number of Items | |
| C+ | 0 | 0 |
| B+ | 3 | 0 |
| A+ | 6 | 8 |
| A– | 6 | 10 |
| B– | 2 | 0 |
| C– | 1 | 0 |

## Grade 11 Writing Form C

Table 01. Grade 11 Writing: Item Analysis Summary Statistics

| N=576 | Item Difficulty | Item Discrimination |
|---|---|---|
| Mean | 0.51 | 0.50 |
| SD | 0.16 | 0.24 |
| Min | 0.07 | -0.22 |
| Max | 0.82 | 0.82 |

Table 02. Grade 11 Writing: Distribution of DIF Item Categorizations for Selected Groups

| | Female vs. Male | Hispanic vs. White |
|---|---|---|
| | Number of Students | |
| Focal | 283 | 118 |
| Reference | 293 | 322 |
| | Number of Items | |
| C+ | 0 | 1 |
| B+ | 4 | 0 |
| A+ | 6 | 8 |
| A– | 6 | 7 |
| B– | 2 | 0 |
| C– | 1 | 3 |

Agreement Tables for Multiple Dimension Items

Grade 4          43504
                 ORGANIZATION

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 227 | 34 | 0 | 261 |
| 2 | 41 | 107 | 4 | 152 |
| 3 | 0 | 18 | 7 | 25 |
| 4 | 0 | 0 | 1 | 1 |
| Total | 268 | 159 | 12 | 439 |

Grade 4          43504
                 ORGANIZATION

| First Human Scorer by Second Human Scorer | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 237 | 23 | 1 | 0 | 261 |
| 2 | 27 | 122 | 3 | 0 | 152 |
| 3 | 0 | 6 | 19 | 0 | 25 |
| 4 | 0 | 0 | 0 | 1 | 1 |
| Total | 264 | 151 | 23 | 1 | 439 |

Grade 4          43504
                 ELABORATION

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 220 | 37 | 0 | 257 |
| 2 | 32 | 124 | 5 | 161 |
| 3 | 0 | 12 | 7 | 19 |
| 4 | 0 | 0 | 2 | 2 |
| Total | 252 | 173 | 14 | 439 |

Grade 4          43504
                 ELABORATION

| First Human Scorer | Second Human Scorer | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Total |
| 1 | 234 | 23 | 0 | 0 | 257 |
| 2 | 26 | 131 | 4 | 0 | 161 |
| 3 | 0 | 6 | 13 | 0 | 19 |
| 4 | 0 | 1 | 0 | 1 | 2 |
| Total | 260 | 161 | 17 | 1 | 439 |

First Human Scorer by Second Human Scorer

Grade 4          43504
                 CONVENTIONS

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 46 | 85 | 4 | 135 |
| 1 | 27 | 103 | 39 | 169 |
| 2 | 5 | 78 | 52 | 135 |
| Total | 78 | 266 | 95 | 439 |

First Human Scorer by Automated Score

115

Grade 4 43334
ORGANIZATION

| First Human Scorer by Automated Score | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Automated Score | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 204 | 17 | 2 | 0 | 223 |
| 2 | 34 | 167 | 13 | 0 | 214 |
| 3 | 0 | 17 | 29 | 1 | 47 |
| 4 | 0 | 0 | 0 | 1 | 1 |
| Total | 238 | 201 | 44 | 2 | 485 |

Grade 4 43334
ORGANIZATION

| First Human Scorer by Second Human Scorer | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 195 | 28 | 0 | 0 | 223 |
| 2 | 30 | 172 | 12 | 0 | 214 |
| 3 | 0 | 11 | 36 | 0 | 47 |
| 4 | 0 | 0 | 0 | 1 | 1 |
| Total | 225 | 211 | 48 | 1 | 485 |

Grade 4 43334
ELABORATION

| First Human Scorer by Automated Score | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Automated Score | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 208 | 16 | 2 | 0 | 226 |
| 2 | 29 | 172 | 14 | 1 | 216 |
| 3 | 0 | 18 | 24 | 1 | 43 |
| Total | 237 | 206 | 40 | 2 | 485 |

Grade 4 43334
ELABORATION

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 196 | 30 | 0 | 226 |
| 2 | 31 | 171 | 14 | 216 |
| 3 | 0 | 11 | 32 | 43 |
| Total | 227 | 212 | 46 | 485 |

Grade 4                43334
                       CONVENTIONS

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 0 | 1 | 2 | Total |
| 0 | 79 | 95 | 7 | 181 |
| 1 | 40 | 119 | 28 | 187 |
| 2 | 3 | 66 | 48 | 117 |
| Total | 122 | 280 | 83 | 485 |

Grade 4

                       43284
                       ORGANIZATION

| First Human Scorer by Automated Score | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Automated Score | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 176 | 36 | 1 | 0 | 213 |
| 2 | 28 | 94 | 10 | 0 | 132 |
| 3 | 1 | 13 | 31 | 2 | 47 |
| 4 | 0 | 0 | 2 | 2 | 4 |
| Total | 205 | 143 | 44 | 4 | 396 |

Grade 4                43284
                       ORGANIZATION

| First Human Scorer by Second Human Scorer | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 181 | 32 | 0 | 0 | 213 |
| 2 | 25 | 102 | 5 | 0 | 132 |
| 3 | 0 | 14 | 29 | 4 | 47 |
| 4 | 0 | 0 | 1 | 3 | 4 |
| Total | 206 | 148 | 35 | 7 | 396 |

Grade 4    43284
           ELABORATION

| First Human Scorer by Automated Score | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Automated Score | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 175 | 38 | 0 | 0 | 213 |
| 2 | 28 | 97 | 7 | 0 | 132 |
| 3 | 1 | 13 | 32 | 2 | 48 |
| 4 | 0 | 0 | 2 | 1 | 3 |
| Total | 204 | 148 | 41 | 3 | 396 |

Grade 4    43284
           ELABORATION

| First Human Scorer by Second Human Scorer | | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 180 | 33 | 0 | 0 | 213 |
| 2 | 23 | 104 | 5 | 0 | 132 |
| 3 | 0 | 13 | 32 | 3 | 48 |
| 4 | 0 | 0 | 1 | 2 | 3 |
| Total | 203 | 150 | 38 | 5 | 396 |

Grade 4    43284
           CONVENTIONS

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 0 | 1 | 2 | Total |
| 0 | 81 | 59 | 11 | 151 |
| 1 | 66 | 74 | 18 | 158 |
| 2 | 16 | 41 | 30 | 87 |
| Total | 163 | 174 | 59 | 396 |

Grade 7          43438
                 ORGANIZATION

| First Human Scorer by Automated Score | | | | |
| --- | --- | --- | --- | --- |
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 63 | 16 | 0 | 79 |
| 2 | 17 | 51 | 5 | 73 |
| 3 | 0 | 5 | 7 | 12 |
| Total | 80 | 72 | 12 | 164 |

Grade 7          43438
                 ORGANIZATION

| First Human Scorer by Second Human Scorer | | | | |
| --- | --- | --- | --- | --- |
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 69 | 10 | 0 | 79 |
| 2 | 13 | 58 | 2 | 73 |
| 3 | 2 | 3 | 7 | 12 |
| Total | 84 | 71 | 9 | 164 |

Grade 7          43438
                 ELABORATION

| First Human Scorer by Automated Score | | | | |
| --- | --- | --- | --- | --- |
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 69 | 12 | 0 | 81 |
| 2 | 15 | 52 | 5 | 72 |
| 3 | 0 | 5 | 6 | 11 |
| Total | 84 | 69 | 11 | 164 |

**Grade 7**  43438  
**ELABORATION**

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 71 | 10 | 0 | 81 |
| 2 | 14 | 56 | 2 | 72 |
| 3 | 2 | 2 | 7 | 11 |
| Total | 87 | 68 | 9 | 164 |

First Human Scorer by Second Human Scorer

**Grade 7**  43438  
**CONVENTIONS**

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 15 | 31 | 4 | 50 |
| 1 | 6 | 32 | 19 | 57 |
| 2 | 3 | 16 | 38 | 57 |
| Total | 24 | 79 | 61 | 164 |

First Human Scorer by Automated Score

**Grade 7**  43703  
**ORGANIZATION**

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 96 | 14 | 0 | 110 |
| 2 | 12 | 48 | 1 | 61 |
| 3 | 0 | 4 | 0 | 4 |
| Total | 108 | 66 | 1 | 175 |

First Human Scorer by Automated Score

Grade 7        43703
**ORGANIZATION**

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 101 | 9 | 0 | 110 |
| 2 | 14 | 43 | 4 | 61 |
| 3 | 0 | 3 | 1 | 4 |
| Total | 115 | 55 | 5 | 175 |

Grade 7        43703
**ELABORATION**

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 96 | 16 | 0 | 112 |
| 2 | 10 | 50 | 1 | 61 |
| 3 | 0 | 2 | 0 | 2 |
| Total | 106 | 68 | 1 | 175 |

43703
Grade 7        **ELABORATION**

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 103 | 9 | 0 | 112 |
| 2 | 12 | 45 | 4 | 61 |
| 3 | 0 | 2 | 0 | 2 |
| Total | 115 | 56 | 4 | 175 |

Grade 7        43703
               **CONVENTIONS**

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | **0** | **1** | **2** | **Total** |
| **0** | 37 | 26 | 0 | 63 |
| **1** | 20 | 28 | 14 | 62 |
| **2** | 7 | 15 | 29 | 51 |
| **Total** | 64 | 69 | 43 | 176 |

*First Human Scorer by Automated Score*

Grade 7        43469
               **ORGANIZATION**

| First Human Scorer | Automated Score | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **Total** |
| **1** | 71 | 7 | 0 | 0 | 78 |
| **2** | 8 | 46 | 4 | 0 | 58 |
| **3** | 0 | 8 | 6 | 0 | 14 |
| **4** | 0 | 0 | 0 | 3 | 3 |
| **Total** | 79 | 61 | 10 | 3 | 153 |

*First Human Scorer by Automated Score*

Grade 7        43469
               **ORGANIZATION**

| First Human Scorer | Second Human Scorer | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **Total** |
| **1** | 70 | 8 | 0 | 0 | 78 |
| **2** | 8 | 45 | 5 | 0 | 58 |
| **3** | 0 | 7 | 7 | 0 | 14 |
| **4** | 0 | 0 | 0 | 3 | 3 |
| **Total** | 78 | 60 | 12 | 3 | 153 |

*First Human Scorer by Second Human Scorer*

Grade 7          43469
                 **ELABORATION**

| First Human Scorer | First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Automated Score | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 75 | 6 | 0 | 0 | 81 |
| 2 | 8 | 44 | 4 | 0 | 56 |
| 3 | 0 | 8 | 5 | 0 | 13 |
| 4 | 0 | 0 | 1 | 2 | 3 |
| Total | 83 | 58 | 10 | 2 | 153 |

Grade 7          43469
                 **ELABORATION**

| First Human Scorer | First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | | |
| | 1 | 2 | 3 | 4 | Total |
| 1 | 72 | 9 | 0 | 0 | 81 |
| 2 | 9 | 42 | 5 | 0 | 56 |
| 3 | 0 | 7 | 6 | 0 | 13 |
| 4 | 0 | 0 | 0 | 3 | 3 |
| Total | 81 | 58 | 11 | 3 | 153 |

Grade 7          43469
                 **CONVENTIONS**

| First Human Scorer | First Human Scorer by Automated Score | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 0 | 1 | 2 | Total |
| 0 | 43 | 13 | 1 | 57 |
| 1 | 23 | 36 | 7 | 66 |
| 2 | 3 | 12 | 18 | 33 |
| Total | 69 | 61 | 26 | 156 |

Grade 11   43632
        ORGANIZATION

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 15 | 3 | 0 | 18 |
| 2 | 2 | 32 | 1 | 35 |
| 3 | 0 | 3 | 8 | 11 |
| Total | 17 | 38 | 9 | 64 |

Grade 11   43632
        ORGANIZATION

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 15 | 3 | 0 | 18 |
| 2 | 1 | 30 | 4 | 35 |
| 3 | 0 | 1 | 10 | 11 |
| Total | 16 | 34 | 14 | 64 |

Grade 11   43632
        ELABORATION

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 15 | 4 | 0 | 19 |
| 2 | 7 | 29 | 0 | 36 |
| 3 | 0 | 0 | 9 | 9 |
| Total | 22 | 33 | 9 | 64 |

Grade 11      43632
**ELABORATION**

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 15 | 4 | 0 | 19 |
| 2 | 2 | 31 | 3 | 36 |
| 3 | 0 | 1 | 8 | 9 |
| Total | 17 | 36 | 11 | 64 |

First Human Scorer by Second Human Scorer

Grade 11      43632
**CONVENTIONS**

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 4 | 2 | 1 | 7 |
| 1 | 2 | 13 | 10 | 25 |
| 2 | 2 | 9 | 21 | 32 |
| Total | 8 | 24 | 32 | 64 |

First Human Scorer by Automated Score

Grade 11      43635
**ORGANIZATION**

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 14 | 1 | 0 | 15 |
| 2 | 3 | 23 | 3 | 29 |
| 3 | 0 | 5 | 16 | 21 |
| Total | 17 | 29 | 19 | 65 |

First Human Scorer by Automated Score

Grade 11      43635
**ORGANIZATION**

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 14 | 1 | 0 | 15 |
| 2 | 3 | 23 | 3 | 29 |
| 3 | 0 | 4 | 17 | 21 |
| Total | 17 | 28 | 20 | 65 |

Grade 11      43635
**ELABORATION**

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 1 | 2 | 3 | Total |
| 1 | 14 | 1 | 0 | 15 |
| 2 | 2 | 27 | 5 | 34 |
| 3 | 0 | 2 | 14 | 16 |
| Total | 16 | 30 | 19 | 65 |

Grade 11      43635
**ELABORATION**

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 14 | 1 | 0 | 15 |
| 2 | 3 | 28 | 3 | 34 |
| 3 | 0 | 1 | 15 | 16 |
| Total | 17 | 30 | 18 | 65 |

Grade 11          43635
                  CONVENTIONS

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 1 | 9 | 1 | 11 |
| 1 | 2 | 16 | 12 | 30 |
| 2 | 0 | 4 | 20 | 24 |
| Total | 3 | 29 | 33 | 65 |

**First Human Scorer by Automated Score**

Grade 11          43479
                  ORGANIZATION

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 20 | 4 | 0 | 24 |
| 2 | 5 | 30 | 3 | 38 |
| 3 | 0 | 1 | 6 | 7 |
| Total | 25 | 35 | 9 | 69 |

**First Human Scorer by Automated Score**

Grade 11          43479
                  ORGANIZATION

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Total |
| 1 | 21 | 3 | 0 | 24 |
| 2 | 1 | 33 | 4 | 38 |
| 3 | 0 | 0 | 7 | 7 |
| Total | 22 | 36 | 11 | 69 |

**First Human Scorer by Second Human Scorer**

Grade 11    43479
ELABORATION

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| First Human Scorer by Automated Score | | | | |
| First Human Scorer | 1 | 2 | 3 | Total |
| 1 | 24 | 4 | 0 | 28 |
| 2 | 8 | 24 | 1 | 33 |
| 3 | 0 | 3 | 5 | 8 |
| Total | 32 | 31 | 6 | 69 |

Grade 11    43479
ELABORATION

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 1 | 2 | 3 | Total |
| 1 | 23 | 5 | 0 | 28 |
| 2 | 3 | 30 | 0 | 33 |
| 3 | 0 | 3 | 5 | 8 |
| Total | 26 | 38 | 5 | 69 |

Grade 11    43479
CONVENTIONS

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 0 | 1 | 2 | Total |
| 0 | 4 | 6 | 0 | 10 |
| 1 | 5 | 15 | 8 | 28 |
| 2 | 1 | 11 | 19 | 31 |
| Total | 10 | 32 | 27 | 69 |

Grade 11          43479
CONVENTIONS

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 0 | 1 | 2 | Total |
| 0 | 8 | 2 | 0 | 10 |
| 1 | 3 | 22 | 3 | 28 |
| 2 | 0 | 9 | 22 | 31 |
| Total | 11 | 33 | 25 | 69 |

Appendix E

Agreement Tables for Propositional Model

Grade 4 MATH          43572

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | |
| | **0** | **1** | **Total** |
| 0 | 288 | 25 | 313 |
| 1 | 115 | 104 | 219 |
| Total | 403 | 129 | 532 |

Grade 4 MATH          43572

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | |
| | **0** | **1** | **Total** |
| 0 | 300 | 13 | 313 |
| 1 | 13 | 206 | 219 |
| Total | 313 | 219 | 532 |

Grade 4 MATH          43564

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | |
| | **0** | **1** | **Total** |
| 0 | 155 | 0 | 155 |
| 1 | 0 | 1 | 1 |
| Total | 155 | 1 | 156 |

Grade 4 MATH        43564

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| First Human Scorer | Second Human Scorer | | |
| | 0 | 1 | Total |
| 0 | 155 | 0 | 155 |
| 1 | 0 | 1 | 1 |
| Total | 155 | 1 | 156 |

Grade 4 MATH        43173

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| First Human Scorer | Automated Score | | | |
| | 0 | 1 | 2 | Total |
| 0 | 447 | 3 | 0 | 450 |
| 1 | 34 | 38 | 0 | 72 |
| 2 | 1 | 0 | 1 | 2 |
| Total | 482 | 41 | 1 | 524 |

Grade 4 MATH        43173

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 0 | 1 | 2 | Total |
| 0 | 446 | 4 | 0 | 450 |
| 1 | 5 | 65 | 2 | 72 |
| 2 | 0 | 1 | 1 | 2 |
| Total | 451 | 70 | 3 | 524 |

Grade 7 MATH        43551

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| First Human Scorer | Automated Score | | |
| | 0 | 1 | Total |
| 0 | 148 | 11 | 159 |
| 1 | 28 | 21 | 49 |

| | | | Total | 176 | 32 | 208 |
|---|---|---|---|---|---|---|

Grade 7 MATH 43551

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| First Human Scorer | Second Human Scorer | | |
| | 0 | 1 | Total |
| 0 | 153 | 6 | 159 |
| 1 | 12 | 37 | 49 |
| Total | 165 | 43 | 208 |

Grade 7 MATH 43555

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| First Human Scorer | Automated Score | | |
| | 0 | 1 | Total |
| 0 | 91 | 22 | 113 |
| 1 | 34 | 43 | 77 |
| 2 | 1 | 1 | 2 |
| Total | 126 | 66 | 192 |

Grade 7 MATH 43555

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 0 | 1 | 2 | Total |
| 0 | 106 | 7 | 0 | 113 |
| 1 | 9 | 68 | 0 | 77 |
| 2 | 0 | 0 | 2 | 2 |
| Total | 115 | 75 | 2 | 192 |

Grade 7 MATH 43557

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| First Human Scorer | Automated Score | | |
| | 0 | 1 | Total |
| 0 | 119 | 9 | 128 |
| 1 | 19 | 42 | 61 |

| Total | 138 | 51 | 189 |
|-------|-----|-----|-----|

Grade 7 MATH          43557

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | |
| | 0 | 1 | Total |
| 0 | 125 | 3 | 128 |
| 1 | 5 | 56 | 61 |
| Total | 130 | 59 | 189 |

Grade 7 MATH          43639

| First Human Scorer by Automated Score | | | |
|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | |
| | 0 | 1 | Total |
| 0 | 115 | 3 | 118 |
| 1 | 18 | 50 | 68 |
| Total | 133 | 53 | 186 |

Grade 7 MATH          43639

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | |
| | 0 | 1 | Total |
| 0 | 114 | 4 | 118 |
| 1 | 5 | 63 | 68 |
| Total | 119 | 67 | 186 |

Grade 7 MATH          43559

| First Human Scorer by Automated Score | | |
|---|---|---|
| First Human Scorer | Automated Score | |
| | 0 | Total |
| 0 | 52 | 52 |
| 1 | 2 | 2 |
| Total | 54 | 54 |

Grade 7 MATH          43559

| First Human Scorer by Second Human Scorer | | | |
|---|---|---|---|
| First Human Scorer | Second Human Scorer | | |
| | 0 | 1 | Total |
| 0 | 52 | 0 | 52 |
| 1 | 1 | 1 | 2 |
| Total | 53 | 1 | 54 |

Grade 7 MATH          43552

| First Human Scorer by Automated Score | | |
|---|---|---|
| First Human Scorer | Automated Score | |
| | 0 | Total |
| 0 | 24 | 24 |
| Total | 24 | 24 |

Grade 7 MATH          43552

| First Human Scorer by Second Human Scorer | | |
|---|---|---|
| First Human Scorer | Second Human Scorer | |
| | 0 | Total |
| 0 | 24 | 24 |
| Total | 24 | 24 |

Grade 4 READING          43707

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 298 | 1 | 2 | 301 |
| 1 | 72 | 51 | 3 | 126 |
| 2 | 60 | 17 | 36 | 113 |
| Total | 430 | 69 | 41 | 540 |

**First Human Scorer by Automated Score**

Grade 4 READING          43707

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 287 | 7 | 7 | 301 |
| 1 | 9 | 104 | 13 | 126 |
| 2 | 4 | 9 | 100 | 113 |
| Total | 300 | 120 | 120 | 540 |

**First Human Scorer by Second Human Scorer**

Grade 4 READING          43412

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 324 | 9 | 0 | 333 |
| 1 | 69 | 89 | 9 | 167 |
| 2 | 3 | 12 | 9 | 24 |
| Total | 396 | 110 | 18 | 524 |

**First Human Scorer by Automated Score**

Grade 4 READING        43412

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | | |
| | **0** | **1** | **2** | **Total** |
| 0 | 315 | 18 | 0 | 333 |
| 1 | 18 | 143 | 6 | 167 |
| 2 | 0 | 8 | 16 | 24 |
| Total | 333 | 169 | 22 | 524 |

Grade 4 READING        43416

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | | |
| | **0** | **1** | **2** | **Total** |
| 0 | 223 | 12 | 1 | 236 |
| 1 | 14 | 181 | 8 | 203 |
| 2 | 2 | 47 | 49 | 98 |
| Total | 239 | 240 | 58 | 537 |

Grade 4 READING        43416

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | | |
| | **0** | **1** | **2** | **Total** |
| 0 | 229 | 6 | 1 | 236 |
| 1 | 7 | 188 | 8 | 203 |
| 2 | 1 | 16 | 81 | 98 |
| Total | 237 | 210 | 90 | 537 |

Grade 7 READING        43248

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | | |
| | **0** | **1** | **2** | **Total** |
| 0 | 108 | 3 | 0 | 111 |
| 1 | 77 | 13 | 0 | 90 |
| 2 | 5 | 1 | 1 | 7 |
| Total | 190 | 17 | 1 | 208 |

Grade 7 READING          43248

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | | |
| | **0** | **1** | **2** | **Total** |
| **0** | 91 | 19 | 1 | 111 |
| **1** | 18 | 66 | 6 | 90 |
| **2** | 0 | 3 | 4 | 7 |
| **Total** | 109 | 88 | 11 | 208 |

Grade 7 READING          43445

| First Human Scorer by Automated Score | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Automated Score** | | | |
| | **0** | **1** | **2** | **Total** |
| **0** | 118 | 5 | 1 | 124 |
| **1** | 31 | 8 | 0 | 39 |
| **2** | 11 | 10 | 2 | 23 |
| **Total** | 160 | 23 | 3 | 186 |

Grade 7 READING          43445

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| **First Human Scorer** | **Second Human Scorer** | | | |
| | **0** | **1** | **2** | **Total** |
| **0** | 110 | 13 | 1 | 124 |
| **1** | 14 | 23 | 2 | 39 |
| **2** | 4 | 4 | 15 | 23 |
| **Total** | 128 | 40 | 18 | 186 |

Grade 7 READING          43422

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 121 | 2 | 0 | 123 |
| 1 | 33 | 6 | 0 | 39 |
| 2 | 18 | 4 | 4 | 26 |
| Total | 172 | 12 | 4 | 188 |

First Human Scorer by Automated Score

Grade 7 READING          43422

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 111 | 7 | 5 | 123 |
| 1 | 9 | 29 | 1 | 39 |
| 2 | 6 | 2 | 18 | 26 |
| Total | 126 | 38 | 24 | 188 |

First Human Scorer by Second Human Scorer

Grade 11 READING          43297

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 31 | 0 | 0 | 31 |
| 1 | 13 | 10 | 0 | 23 |
| 2 | 1 | 3 | 2 | 6 |
| Total | 45 | 13 | 2 | 60 |

First Human Scorer by Automated Score

Grade 11 READING          43297

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 29 | 2 | 0 | 31 |
| 1 | 4 | 16 | 3 | 23 |
| 2 | 0 | 1 | 5 | 6 |
| Total | 33 | 19 | 8 | 60 |

Grade 11 READING          43435

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 40 | 2 | 0 | 42 |
| 1 | 12 | 4 | 2 | 18 |
| 2 | 2 | 1 | 1 | 4 |
| Total | 54 | 7 | 3 | 64 |

Grade 11 READING          43435

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 40 | 2 | 0 | 42 |
| 1 | 7 | 11 | 0 | 18 |
| 2 | 2 | 1 | 1 | 4 |
| Total | 49 | 14 | 1 | 64 |

Grade 11 READING          43397

| First Human Scorer | Automated Score | | |
| --- | --- | --- | --- |
| | 0 | 1 | Total |
| 0 | 29 | 2 | 31 |
| 1 | 1 | 2 | 3 |
| 2 | 0 | 5 | 5 |
| Total | 30 | 9 | 39 |

Table title: First Human Scorer by Automated Score

Grade 11 READING          43397

| First Human Scorer | Second Human Scorer | | | |
| --- | --- | --- | --- | --- |
| | 0 | 1 | 2 | Total |
| 0 | 31 | 0 | 0 | 31 |
| 1 | 0 | 3 | 0 | 3 |
| 2 | 1 | 0 | 4 | 5 |
| Total | 32 | 3 | 4 | 39 |

Table title: First Human Scorer by Second Human Scorer

Grade 4 WRITING          43280
Form C

| First Human Scorer | Automated Score | | | |
| --- | --- | --- | --- | --- |
| | 0 | 1 | 2 | Total |
| 0 | 345 | 14 | 0 | 359 |
| 1 | 50 | 39 | 2 | 91 |
| 2 | 9 | 17 | 9 | 35 |
| Total | 404 | 70 | 11 | 485 |

Table title: First Human Scorer by Automated Score

Grade 4 WRITING    43280
Form C

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 339 | 20 | 0 | 359 |
| 1 | 11 | 79 | 1 | 91 |
| 2 | 1 | 7 | 27 | 35 |
| Total | 351 | 106 | 28 | 485 |

First Human Scorer by Second Human Scorer

Grade 7 WRITING    43468
Form C

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 70 | 5 | 2 | 77 |
| 1 | 9 | 30 | 6 | 45 |
| 2 | 1 | 7 | 35 | 43 |
| Total | 80 | 42 | 43 | 165 |

First Human Scorer by Automated Score

Grade 7 WRITING    43468
Form C

| First Human Scorer | Second Human Scorer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 77 | 0 | 0 | 77 |
| 1 | 2 | 39 | 4 | 45 |
| 2 | 0 | 0 | 43 | 43 |
| Total | 79 | 39 | 47 | 165 |

First Human Scorer by Second Human Scorer

Grade 11 WRITING    43491
Form C

| First Human Scorer | Automated Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | Total |
| 0 | 9 | 2 | 0 | 11 |
| 1 | 13 | 10 | 1 | 24 |
| 2 | 6 | 6 | 8 | 20 |
| Total | 28 | 18 | 9 | 55 |

First Human Scorer by Automated Score

Grade 11 WRITING 43491
Form C

| First Human Scorer by Second Human Scorer | | | | |
|---|---|---|---|---|
| First Human Scorer | Second Human Scorer | | | |
| | 0 | 1 | 2 | Total |
| 0 | 10 | 1 | 0 | 11 |
| 1 | 2 | 20 | 2 | 24 |
| 2 | 0 | 1 | 19 | 20 |
| Total | 12 | 22 | 21 | 55 |

Appendix I –    Adaptive Item Selection Report

# Smarter Balanced Adaptive Item Selection Algorithm Design Report

Preview Release 16 May 2014

**Jon Cohen and Larry Albright**
**American Institutes for Research**

Produced for Smarter Balanced by the American Institutes for Research

## TABLE OF CONTENTS

# SMARTER BALANCED ADAPTIVE ITEM SELECTION ALGORITHM

## 1. INTRODUCTION, BACKGROUND, AND DEFINITIONS

This document describes the Smarter Balanced adaptive item selection algorithm. The item selection algorithm is designed to cover a standards-based blueprint, which may include content, cognitive complexity, and item type constraints. The item selection algorithm will also include:

- the ability to customize an item pool based on access constraints and screen items that have been previously viewed or may not be accessible for a given individual;

- a mechanism for inserting embedded field-test items; and

- a mechanism for delivering "segmented" tests in which separate parts of the test are administered in a fixed order.

This document describes the algorithm and the design for its implementation for the Smarter Balanced Test Delivery System. The implementation builds extensively on the algorithm implemented in AIR's Test Delivery System. The implementation described is released under a Creative Commons Attribution, No Derivatives license.

The general approach described here is based on a highly parameterized multiple-objective utility function. The objective function includes:

- a measure of content match to the blueprint;

- a measure of overall test information; and

- measures of test information for each reporting category on the test.

We define an objective function that measures an item's contribution to each of these objectives, weighting them to achieve the desired balance among them. Equation 1 sketches this objective function for a single item.

$$f_{ijt} = w_2 \frac{1}{\sum_{r=1}^{R} d_{rj}} \sum_{r=1}^{R} s_{rit} p_r d_{rj} + w_1 \sum_{k=1}^{K} q_k h_{1k}(v_{kijt}, V_{kit}, t_k) + w_0 h_0(u_{ijt}, U_{it}, t_0) \tag{1}$$

where the terms $w$ represent user-supplied weights that assign relative importance to meeting each of the objectives, $d_{rj}$ indicates whether item $j$ has the blueprint-specified feature $r$, and $p_r$ is the user-supplied priority weight for feature $r$. The term $s_{rit}$ is an adaptive control parameter that is described below. In general, $s_{rit}$ increases for features that have not met their designated minimum as the end of the test approaches.

The remainder of the terms represents an item's contribution to measurement precision:

- $v_{kijt}$ is the value of item $j$ toward reducing the measurement error for reporting category $k$ for examinee $i$ at selection $t$; and

- $u_{ijt}$ is the value of item $j$ in terms of reducing the overall measurement error for examinee $i$ at selection $t$.

The terms $U_{it}$ and $V_{kit}$ represent the total information overall and on reporting category $k$, respectively.

The term $q_k$ is a user-supplied priority weight associated with the precision of the score estimate for reporting category $k$. The terms $t$ represent precision targets for the overall score ($t_0$) and each score reporting category score. The functions *h(.)* are given by:

$$h_0\big(u_{ijt}, U_{it}, t_0\big) = \begin{cases} au_{ijt} \text{ if } U_{it} < t_0 \\ bu_{ijt} \text{ otherwise} \end{cases}$$

$$h_{1k}\big(v_{kijt}, V_{kit}, t_k\big) = \begin{cases} c_k v_{kijt} \text{ if } V_{kit} < t_k \\ d_k v_{kijt} \text{ otherwise} \end{cases}$$

Items can be selected to maximize the value of this function. This objective function can be manipulated to produce a pure, standards-free adaptive algorithm by setting $w_2$ to zero or a completely blueprint-driven test by setting $w_1 = w_0 = 0$. Adjusting the weights to optimize performance for a given item pool will enable users to maximize information subject to the constraint that the blueprint is virtually always met.

We note that the computations of the content values and information values generate values on very different scales and that the scale of the content value varies as the test progresses. Therefore, we normalize both the information and content values before computing the value of Equation 1. This normalization is given by $x = \begin{cases} 1 \text{ if } min = max \\ \frac{v - min}{max - min} \text{ otherwise} \end{cases}$, where *min* and *max* represent the minimum and maximum, respectively, of the metric computed over the current set of items or item groups.

The remainder of this section describes the overall program flow, the form of the blueprint, and the various value calculations employed in the objective function. Subsequent sections describe the details of the selection algorithm.

## 1.1 Blueprint

Each test will be described by a single blueprint for each segment of the test and will identify the order in which the segments appear. The blueprint will include:

- an indicator of whether the test is adaptive or fixed form;
- termination conditions for the segment, which are described in a subsequent section;
- a set of nested content constraints, each of which is expressed as:
  - the minimum number of items to be administered within the content category;
  - the maximum number of items to be administered within the content category;

- – an indication of whether the maximum should be deterministically enforced (a "strict" maximum);

- – a priority weight for the content category $p_r$;

- – an explicit indicator as to whether this content category is a reporting category; and

- – an explicit precision-priority weight ($q_k$) for each group identified as a reporting category.

- ▪ a set of non-nested content constraints, which are represented as:

- – a name for the collection of items meeting the constraint;

- – the minimum number of items to be administered from this group of items;

- – the maximum number of items to be administered from this group of items;

- – an indication of whether the maximum should be deterministically enforced (a "strict" maximum);

- – a priority weight for the group of items $p_r$;

- – an explicit indicator as to whether this named group will make up a reporting category; and

- – an explicit precision-priority weight ($q_k$) for each group identified as a reporting category.

- – The priority weights, $p_r$ on the blueprint, can be used to express values in the blueprint match. Large weights on reporting categories paired with low (or zero) weights on the content categories below them may allow more flexibility to maximize information in a content category covering fewer fine-grained targets, while the reverse would mitigate toward more reliable coverage of finer-grained categories, with less content flexibility within reporting categories.

An example of a blueprint specification appears in Appendix 1.

Each segment of a test will have a separate blueprint.

## 1.2  Content Value

Each item or item group will be characterized by its contribution to meeting the blueprint, given the items that have already been administered at any point. The contribution is based on the presence or absence of features specified in the blueprint and denoted by the term *d* in Equation 1. This section describes the computation of the content value.

### 1.2.1  Content Value for Single Items

For each constraint appearing in the blueprint (*r*), an item *i* either does or does not have the characteristic described by the constraint. For example, a constraint might require a minimum of four and a maximum of six algebra items. An item measuring algebra has the described

characteristic, and an item measuring geometry but algebra does not. To capture this constraint, we define the following:

- $d_i$ is a feature vector in which the elements are $d_{ir}$, summarizing item $i$'s contribution to meeting the blueprint. This feature vector includes content categories such as claims and targets as well as other features of the blueprint, such as Depth of Knowledge and item type.

- $S_{it}$ is a diagonal matrix, the diagonal elements of which are the adaptive control parameters $s_{rit}$.

- $p$ is the vector containing the user-supplied priority weights $p_r$.

The scalar content value for an item is given by $C_{ijt} = d_i' S_{it} p$.

Letting $z_{rit}$ represent the number of items with feature $r$ administered to student $i$ by iteration $t$, the value of the adaptive control parameters is:

$$
s_{rit} = \begin{cases} m_{it}\left(2 - \dfrac{z_{rit}}{Min_r}\right) & if\ z_r < Min_r \\[2mm] 1 - \dfrac{z_{rit} - Min_r}{Max_r - Min_r} & if\ Min_r < z_{rit} < Max_r \\[2mm] (Max_r - z_{rit}) - 1 & if\ Max_r \leq z_{rit} \end{cases}
$$

The blueprint defines the minimum ($Min_r$) and maximum ($Max_r$) number of items to be administered with each characteristic ($r$).

The term $m_{it} = \dfrac{T}{T - t}$ where $T$ is the total test length. This has the effect of increasing the algorithm's preference for items that have not yet met their minimums as the end of the test nears and the opportunities to meet the minimum diminish.

This increases the likelihood of selecting items for content that has not met its minimum as the opportunities to do so are used up. The value $s$ is highest for items with content that has not met its minimum, declines for items representing content for which the minimum number of items has been reached but the maximum has not, and turns negative for items representing content that has met the maximum.

### 1.2.2 Content Value for Sets of Items

Calculation of the content value of sets of items is complicated by two factors:

1. The desire to allow more items to be developed for each set and to have the most advantageous set of items administered

2. The design objective of characterizing the information contribution of a set of items as the expected information over the working theta distribution for the examinee

The former objective is believed to enhance the ability to satisfy highly constrained blueprints while still adapting to obtain good measurement for a broad range of students. The latter arises from the recognition that ELA tests will select one set of items at a time, without an opportunity to adapt once the passage has been selected.

The general approach involves successive selection of the highest content value item in the set until the indicated number of items in the set have been selected. Because the content value of an item changes with each selection, a temporary copy of the already-administered content vector for the examinee is updated with each selection such that subsequent selections reflect the items selected in previous iterations.

Exhibit 1 presents a flowchart for this calculation. Readers will note the check to determine whether $w_0 > 0$ or $w_1 > 0$. These weights, defined with Equation 1, identify the user-supplied importance of information optimization relative to blueprint optimization. In cases such as independent field tests, this weight may be set to zero, as it may not be desirable to make item administration dependent on match to student performance. In more typical adaptive cases where item statistics will not be recalculated, favoring more informative items is generally better. The final measure of content value for the set of selected set of items is divided by the number of items selected to avoid a bias toward selection of sets with more items.
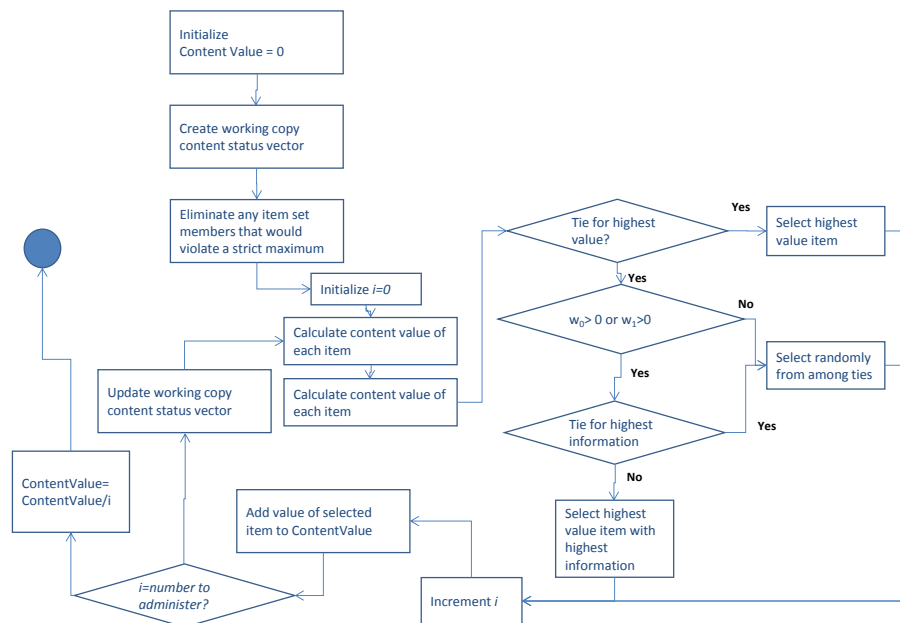
Exhibit 1. Content Value Calculation for Item Sets

## 1.3 Information Value

Each item or item group also has value in terms of maximizing information, both overall and on reporting categories.

### 1.3.1 Individual Information Value

The information value associated with an item will be an approximation of information. The system will be designed to use generalized IRT models; however, it will treat all items as though they offer equal measurement precision. This is the assumption made by the Rasch model, but in more general models, items known to offer better measurement are given preference by many algorithms. Subsequent algorithms are then required to control the exposure of the items that measure best. Ignoring the differences in slopes serves to eliminate this bias and help equalize exposure.

### 1.3.2 Binary Items

The approximate information value of a binary item will be characterized as $I_j(\theta) = p_j(\theta)(1 - p_j(\theta))$, where the slope parameters are artificially replaced with a constant.

### 1.3.3 Polytomous Items

In terms of information, the best polytomous item in the pool is the one that maximizes the expected information, $I_j(\theta)$. Formally, $I_j(\theta) > I_k(\theta)$ for all items $k \neq j$. The true value $\theta$, however, remains unknown and is accessed only through an estimate, $\hat{\theta} \sim N(\bar{\theta}, \sigma_\theta)$. By definition of an expectation, the expected information $I_j(\theta) = \int I_j(t) f(t|\bar{\theta}, \sigma_\theta) dt$.

The intuition behind this result is illustrated in Exhibit 2. In Exhibit 2, each panel graphs the distribution of the estimate of $\theta$ for an examinee. The top panel assumes a polytomous item in which one step threshold (A1) matches the mean of the $\theta$ estimate distribution. In the bottom panel, neither step threshold matches the mean of the $\theta$ estimate distribution. The shaded area in each panel indicates the region in which the hypothetical item depicted in the panel provides more information. We see that approximately 2/3 of the probability density function is shaded in the lower panel, while the item depicted in the upper panel dominates in only about 1/3 of the cases. In this example, the item depicted in the lower panel has a much greater probability of maximizing the information from the item, despite the fact that the item in the upper panel has a threshold exactly matching the mean of the estimate distribution and the item in the lower panel does not.

Threshold A1 matches the best current estimate of the proficiency for this student, but the estimate is not yet very precise

Neither threshold matches the best current estimate of the proficiency for this student, but together they cover more of the proficiency distribution
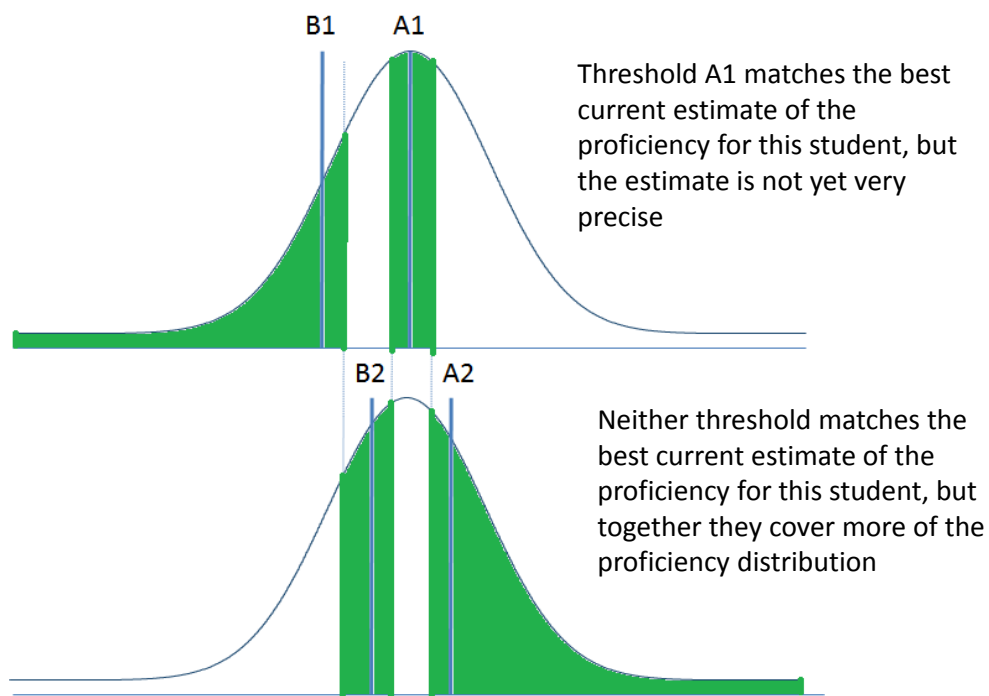
Exhibit 2. Two example items, with the shaded region showing the probability that the item maximizes information for the examinee depicted.

Exhibit 3 shows what happens to information as the estimate of this student's proficiency becomes more precise (later in the test). In this case, the item depicted in the top panel maximizes information about 65-70 percent of the time, compared to about 30 to 35 percent for the item depicted in the lower panel. These are the same items depicted in the Exhibit 2, but in this case we are considering information for a student with a more precise current proficiency estimate.

When the proficiency estimate gets more precise, the item that best matches the center of the distribution covers most of it

As the proficiency distribution becomes more narrow, the item that does not match the center provides less information
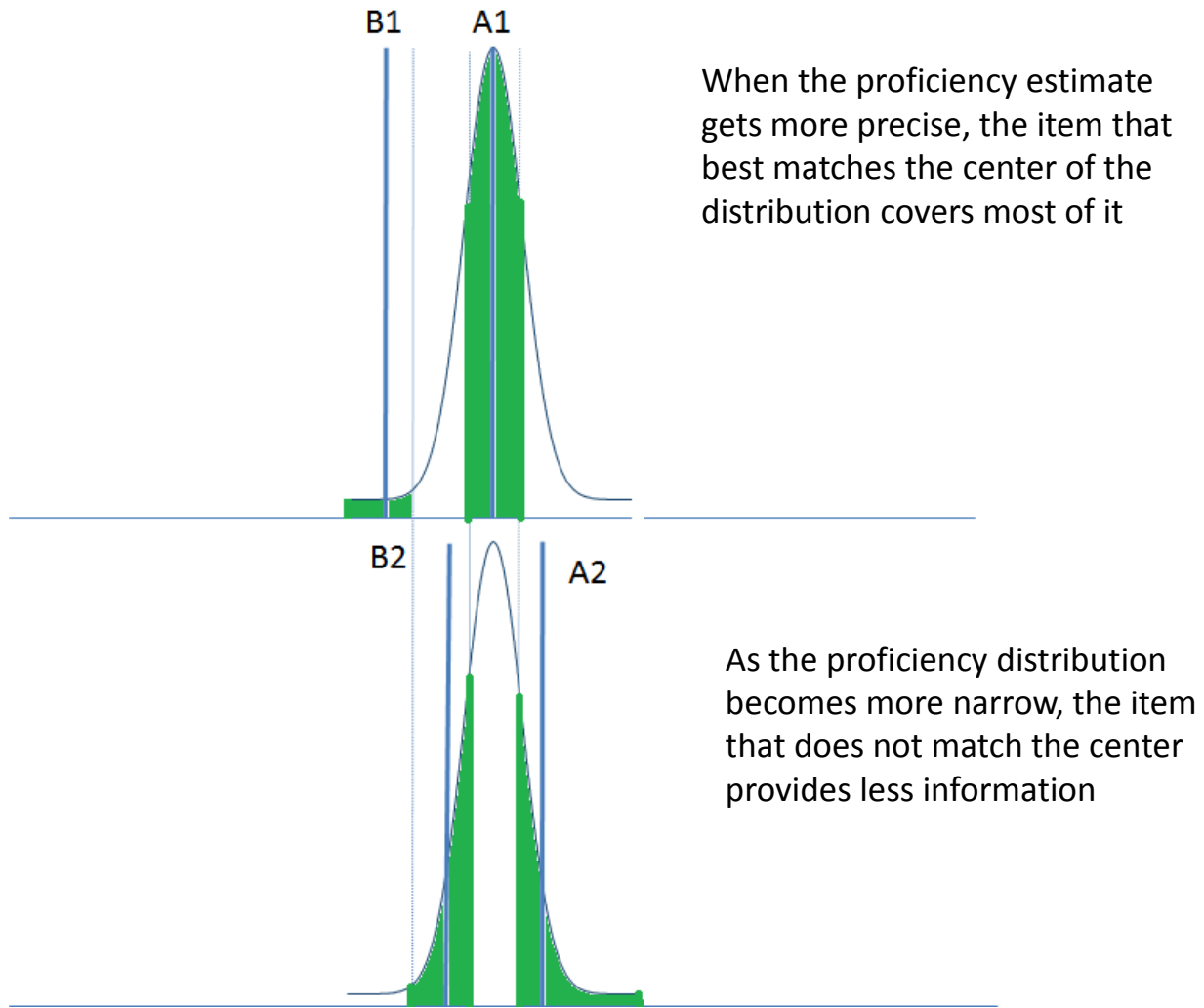
**Exhibit 3: Two example items, with the shaded region showing the probability that the item maximizes information for the examinee depicted.**

The approximate information value of polytomous items will be characterized as the expected information, specifically $E\left[I_j(\theta)|m_i, s_i\right] = \int \sum_{k=1}^{K} I_{jk}(t)\, p_j(k|t)\phi(t; m_i, s_i)dt,$ where $I_{jk}(t)$ represents the information at $t$ of response $k$ to item $j$, $p_j(k|t)$ is the probability of response $k$ to item $j$ (artificially holding slope constant), given proficiency $t$, $\phi(.)$ represents the normal probability density function, and $m_i$ and $s_i$ represent the mean and standard deviation of examinee $i$'s current estimated proficiency distribution.

We propose to use Gauss-Hermite quadrature with a small number of quadrature points (approximately five). Experiments show that we can complete this calculation for 1,000 items in fewer than 5 milliseconds, making it computationally reasonable.

As with the binary items, we propose to ignore the slope parameters to even exposure and avoid a bias toward the items with better measurement.

## 1.3.4   Item Group Information Value

Item groups differ from individual items in that a set of items will be selected for administration. Therefore, the goal is to maximize information across the working theta distribution. As with the polytomous items, we propose to use Gauss-Hermite quadrature to estimate the expected information of the item group.

In the case of multiple-item groups

$$E[I_g(\theta)|m_i, s_i] = \frac{1}{J_g} \int \sum_{j=1}^{J_g} I_{g(j)}(t)\,\phi(t; m_i, s_i)dt$$

Where $I_g(.)$ is the information from item group $g$, $I_{g(j)}$ is the information associated with item $j \in g$, for the $J_g$ items in set $g$. In the case of polytomous items, we use the expected information, as described above.

## 2.   ENTRY AND INITIALIZATION

At startup, the system will

- create a custom item pool;
- initialize theta estimates for the overall score and each score point; and
- insert embedded field-test items.

## 2.1   Item Pool

At test startup the system will generate a *custom item pool*, a string of item IDs for which the student is eligible. This item pool will include all items that

- are active in the system at test startup; and
- are not flagged as "access limited" for attributes associated with this student.

The list will be stored in ascending order of ID.

## 2.2   Adjust Segment Length

Custom item pools run the risk of being unable to meet segment blueprint minimums. To address this special case, the algorithm will adjust the blueprint to be consistent with the custom item pool. This capability becomes necessary when an accommodated item pool systematically excludes some content.

Let

*S* be the set of top-level content constraints in the hierarchical set of constraints, each consisting of the tuple *(name, min, max, n);*

***C*** be the custom item pool, each element consisting of a set of content constraints ***B***;

***f, p*** integers represent item shortfall and pool count, respectively; and

***t*** be the minimum required items on the segment.

For each *s in **S**, c*ompute *n* as the sum of active operational items in ***C*** classified on the constraint.

***f*** = summation over *S* (*min – n*)

***p*** = summation over *S* (*n*)

if *t – f < p*, then *t = t – f*

## 2.3   Initialization of Starting Theta Estimates

The user will supply five pieces of information in the test configuration:

1. A default starting value if no other information is available

2. An indication whether prior scores on the same test should be used, if available

3. Optionally, the test ID of another test that can supply a starting value, along with

4. Slope and intercept parameters to adjust the scale of the value to transform it to the scale of the target test

5. A constant prior variance for use in calculation of working EAP scores

## 2.4   Insertion of Embedded Field-Test Items

Each blueprint will specify

- the number of field-test items to be administered on each test;
- the first item position into which a field-test item may be inserted; and
- the last item position into which a field-test item may be inserted.

Upon startup, select randomly from among the field-test items or item sets until the system has selected the specified number of field-test items. If the items are in sets, the sets will be administered as a complete set, and this may lead to more than the specified number of items administered.

The probability of selection will be given by $p_j = \frac{\sum_{j=1}^{K} K_j}{\sum_{j=1}^{K} a_j K_j} a_j K_j \frac{m}{N_J}$, where

$p_j$ represents the probability of selecting the item;

$m$ is the targeted number of field-test items;

$N_j$ is the total number of active items in the field-test pool;

$K_j$ is the number of items in item set $j$; and

$a_j$ is a user-supplied weight associated with each item (or item set) to adjust the relative probability of selection.

The $a_j$ variables are included to allow for operational cases in which some items must complete field-testing sooner, or enter field-testing later. While using this parameter presents some statistical risk, not doing so poses operational risks.

For each item set, generate a uniform random number $r_j$ on the interval $\{0,1\}$. Sort the items in ascending order by $\frac{r_j}{p_j}$. Sequentially select items, summing the number of items in the set. Stop the selection of field-test items once $FTNMin \leq m \leq FTNMax = \sum_{j=0} K_j$.

Next, each item is assigned to a position on the test. To do so, select a starting position within $f - FTMax - FTMin$ positions from $FTMin$, where $FTMax$ is the maximum allowable position for field-test items and $FTMin$ is the minimum allowable position for field-test items. $FTNMin$ and $FTNMax$ refer to the minimum and maximum number of field-test items, respectively. Distribute the items evenly within these positions.

## 3. ITEM SELECTION

Exhibit 3 summarizes the item selection process. If the item position has been designated for a field-test item, administer that item. Otherwise, the adaptive algorithm kicks in.
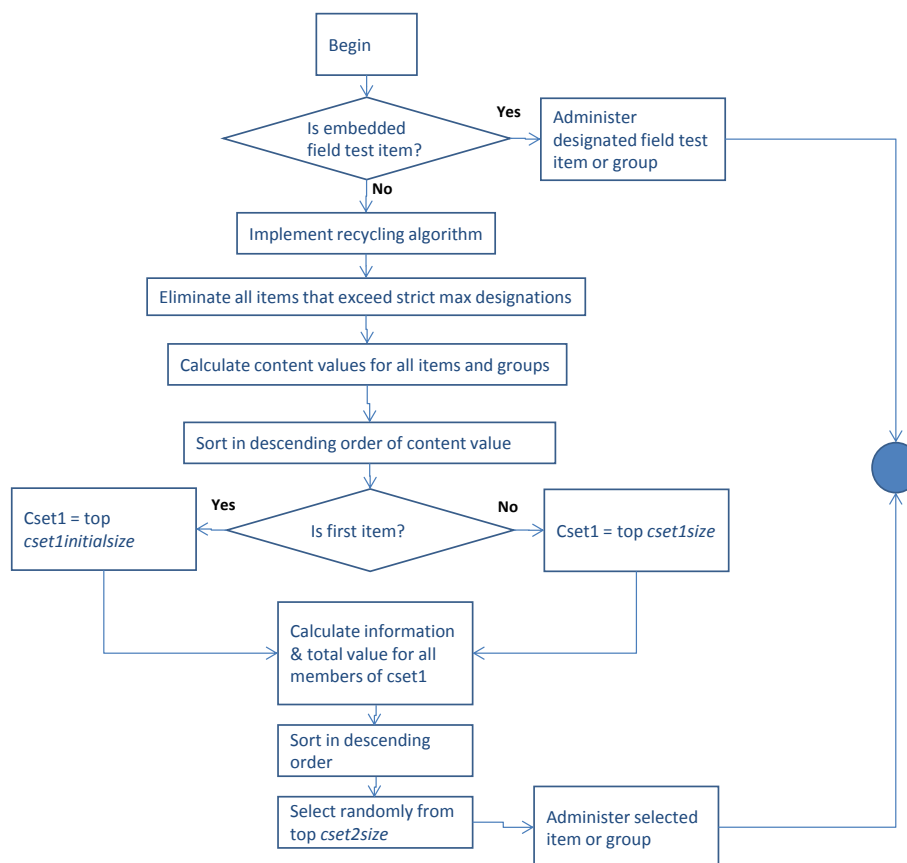
Exhibit 3: Summary of Item Selection Process

This approach is a "content first" approach designed to optimize match to blueprint. An alternative, "information first" approach, is possible. Under an information first approach, all items within a specified information range would be selected as the first set of candidates, and subsequent selection within that set would be based, in part, on content considerations. The engine is being designed so that future development could build such an algorithm using many of the calculations already available.

## 3.1   Trimming the Custom Item Pool

At each item selection, the active item pool is modified in four steps:

1. The custom item pool is intersected with the active item pool, resulting in a custom active item pool.

2. Items already administered on this test are removed from the custom active item pool.

3. Items that have been administered on prior tests are tentatively removed (see Section 3.2 below).

4. Items that measure content that has already exceeded a strict maximum are tentatively removed from the pool, removing entire sets containing items that meet this criterion.

## 3.2   Recycling Algorithm

When students are offered multiple opportunities to test, or when prior tests have been started and invalidated, students will have seen some of the items in the pool. The trimming of the item pool eliminates these items from the pool. It is possible that in such situations, the pool may no longer contain enough items to meet the blueprint.

Hence, items that have been seen on previous administrations may be returned to the pool. If there are not enough items remaining in the pool, the algorithm will recycle items (or item groups) with the required characteristic that is found in insufficient numbers. Working from the least recently administered group, items (or item groups) are reintroduced into the pool until the number of items with the required characteristics meets the minimum requirement. When item groups are recycled, the entire group is recycled rather than an individual item. Items administered on the current test are never recycled.

## 3.3   Adaptive Item Selection

Selection of items will follow a common logic, whether the selection is for a single item or an item group. Item selection will proceed in the following three steps:

1. Select Candidate Set 1 (*cset1*).

   a. Calculate the content value of each item or item group.

   b. Sort the item groups in descending order of content value.

   c. Select the top *cset1size*, a user-supplied value that may vary by test.

2. Select Candidate Set 2 (*cset2*).

   a. Calculate the information values for each item group in *cset1*.

   b. Calculate the overall value of each item group in *cset1* as defined in Equation 1.

   c. Sort *cset2* in descending order of value.

   d. Select the top *cset2size* item groups, where *cset2size* is a user-supplied value that may vary by test.

3. Select the item or item group to be administered.

   a. Select randomly from *cset2* with uniform probability.

Note that a "pure adaptive" test, without regard to content constraints, can be achieved by setting *cset1size* to the size of the item pool and $w_2$, the weight associated meeting content constraints in Equation 1, to zero. Similarly, linear-on-the-fly tests can be constructed by setting $w_0$ and $w_1$ to zero.

## 3.4 Selection of the Initial Item

Selection of the initial item can affect item exposure. At the start of the test, all tests have no content already administered, so the items and item groups have the same content value for all examinees. In general, it is a good idea to spread the initial item selection over a wider range of content values. Therefore, we define an additional user-settable value, *cset1initialsize*, which is the size of Candidate Set 1 on the first item only. Similarly, we define *cset2initialisize*.

## 3.5 Exposure Control

This algorithm uses randomization to control exposure and offers several parameters that can be adjusted to control the tradeoff between optimal item allocation and exposure control. The primary mechanism for controlling exposure is the random selection from *CSET2*, the set of items or item groups that best meet the content and information criteria. These represent the "top *k*" items, where *k* can be set. Larger values of *k* provide more exposure control at the expense of optional selection.

In addition to this mechanism, we avoid a bias toward items with higher measurement precision by treating all items as though they measured with equal precision by ignoring variation in the slope parameter. This has the effect of randomizing over items with differing slope parameters. Without this step, it would be necessary to have other *post hoc* explicit controls to avoid the overexposure of items with higher slope parameters, an approach that could lead to different test characteristics over the course of the testing window.

## 4. TERMINATION

The algorithm will have configurable termination conditions. These may include

- administering a minimum number of items in each reporting category and overall;
- achieving a target level of precision on the overall test score;
- achieving a target level of precision on all reporting categories.

We will define four user-defined flags indicating whether each of these is to be considered in the termination conditions (*TermCount, TermOverall, TermReporting,TermTooClose*). A fifth user-supplied value will indicate whether these are taken in conjunction or if satisfaction of any one of them will suffice (*TermAnd*). Reaching the minimum number of items is always a necessary condition for termination.

In addition, two conditions will each individually and independently cause termination of the test:

1. Administering the maximum number of items specified in the blueprint
2. Having no items in the pool left to administer

## A1. Definitions of User-Settable Parameters

This appendix summarizes the user-settable parameters in the adaptive algorithm.

| Parameter Name | Description | Entity Referred to by Subscript Index |
|---|---|---|
| $w_0$ | Priority weight associated with match to blueprint | N/A |
| $w_1$ | Priority weight associated with reporting category information | N/A |
| $w_2$ | Priority weight associated with overall information | N/A |
| $q_k$ | Priority weight associated with a specific reporting category | reporting categories |
| $p_r$ | Priority weight associated with a feature specified in the blueprint (These inputs appear as a component of the blueprint.) | features specified in the blueprint |
| a | Parameter of the function *h(.)* that controls the overall information weight when the information target has not yet been hit | N/A |
| b | Parameter of the function *h(.)* that controls the overall information weight after the information target has been hit | N/A |
| $c_k$ | Parameter of the function *h(.)* that controls the information weight when the information target has not yet been hit for reporting category *k* | reporting categories |
| $d_k$ | Parameter of the function *h(.)* that controls the information weight after the information target has been hit for reporting category *k* | reporting categories |
| cset1size | Size of candidate pool based on contribution to blueprint match | N/A |
| cset1initialsize | Size of candidate pool based on contribution to blueprint match for the first item or item set selected | N/A |
| cset2size | Size of final candidate pool from which to select randomly | N/A |
| cset2initialsize | Size of candidate pool based on contribution to blueprint match and information for the first item or item set selected | |
| $t_0$ | Target information for the overall test | N/A |
| $t_k$ | Target information for reporting categories | reporting categories |
| startTheta | A default starting value if no other information is available | N/A |
| startPrevious | An indication of whether previous scores on the same test should be used, if available | N/A |
| startOther | The test ID of another test that can supply a starting value, along with startOtherSlope | N/A |
| startOtherSlope | Slope parameter to adjust the scale of the value to transform it to the scale of the target test | N/A |
| startOtherInt | Intercept parameter to adjust the scale of the value to transform it to the scale of the target test | N/A |
| FTMin | Minimum position in which field-test items are allowed | N/A |
| FTMax | Maximum position in which field-test items are allowed | N/A |
| FTNMin | Target minimum number of field-test items | N/A |
| FTNMax | Target maximum number of field-test items | N/A |
| $a_j$ | Weight adjustment for individual embedded field-test items used to increase or decrease their probability of selection | field-test items |

| Parameter Name | Description | Entity Referred to by Subscript Index |
|---|---|---|
| AdaptiveCut | The overall score cutscore, usually proficiency, used in consideration of *TermTooClose* | |
| TooCloseSEs | The number of standard errors below which the difference is considered "too close" to the adaptive cut to proceed. In general, this will signal proceeding to a final segment that contains off-grade items. Ugh. | |
| TermOverall | Flag indicating whether to use the overall information target as a termination criterion | N/A |
| TermReporting | Flag to indicate whether to use reporting category information target as a termination criterion | N/A |
| TermCount | Flag to indicate whether to use minimum test size as a termination condition | N/A |
| TermTooClose | Terminate if you are not sufficiently distant from the specified adaptive cut | |
| TermAnd | Flag to indicate whether the other termination conditions are to be taken separately or conjunctively | N/A |

# A2. API

*This information is forthcoming.*

# A3. Supporting Data Structures

### AIR Cautions and Caveats

- Use of standard error termination conditions will likely cause inconsistencies between the blueprint content specifications and the information criteria will cause unpredictable results, likely leading to failures to meet blueprint requirements.

- The field-test positioning algorithm outlined here is very simple and will lead to deterministic placement of field-test items.

Appendix J –      2013 Pilot Test Report

**From:**     ETS, Smarter Balanced Contract 05: Psychometric Services

**To:**     Smarter Balanced TD&V Leadership Team and Lead Psychometrician

**Subject:**     Pilot Test Data Analysis Results: Dimensionality Study and IRT Model Comparison

**Date:**     April 23, 2014

## Executive Summary

This memorandum contains the statistical analysis summary of results pertaining to the Smarter Balanced Pilot Test, focusing on dimensionality and IRT model investigation. Details related to data processing, student samples, item analysis, DIF analysis, and other relevant factors were summarized in a memorandum to Smarter Balanced in October 2013.

The purposes and design of the Pilot Test administration were documented in the *Pilot Test and Vertical Design* (Educational Testing Service, 2013). The related data collection design is being repeated here to make the conclusions presented in this report more accessible. Caution should be exercised when interpreting information from the Pilot Test administration because of the following limitations:

- The Pilot Test administration used a preliminary version of the Smarter Balanced test blueprints.
- While Pilot tests were being delivered or scored, some items and item types were eliminated.
- Although the initial design was intended to be a representative student sample, the student samples were largely convenience samples.
- The performance task component underwent significant revision after the Pilot Test so that the classroom activity will be a required component of the test administration. Classroom activity was not required during the Pilot.
- The number of scorable performance tasks was very small for some tests, and there were no surviving performance tasks for the mathematics tests.
- Human scoring was performed on the basis of each item, and each item received a maximum of 1,800 scored responses and as few as 500 scored responses for some item types. As a result, not all student responses were fully scored and the sparseness of the analysis data matrix was significant.
- The content data review of the Pilot data for the items was not completed prior to the completion of the analysis activities. Based on preliminary data review, recommendations were implemented concerning which items to include or exclude from the item bank. Items were included if they were not rejected by data review and if they had an item-total correlation no less than 0.15.

The major Pilot statistical analysis activities are item and DIF analyses for CAT items to support data review (completed in October 2013), a dimensionality study to explore grade-level and adjacent-grade dimensional structure, and IRT analyses to provide a basis for the selection of an IRT model.

In the Pilot, students took either a CAT test or a combined CAT and performance task (PT) configuration. Students taking only CAT components took two (ELA) or three (Math) content representative item collections (called modules). Each Math module had 23 selected-response (SR) items and constructed-response (CR) items and was expected to require about 45 minutes to complete. An ELA module had about 29 items at

lower grade levels and 33 items at high school grades and each module was expected to take about 60–75 minutes to complete. All single-selection Selected-Response (SR) items had four choices and multiple-selection Selected-Response (MSR) items had 5 to 8 choices. The performance task items had maximum scores ranging from 0 to 4. In accordance with the test design, other groups of students were administered a single CAT module and a performance task. A performance task was expected to have approximately five scoreable units yielding approximately 20 score points in total. Overall, 1,602 ELA CAT items, 49 ELA performance tasks (which included 318 items), and 1,883 Math CAT items were analyzed. No Math performance tasks were scored and used for any subsequent analyses. These items, in aggregate, represent the ELA and Math in all Claims.

The majority of the Pilot Test contents (CAT modules and PTs) were administered to students at the grade for which the items/tasks were developed (i.e., the on-grade administration of items/tasks). Selected Pilot CAT modules and PTs were also administered to students at the upper- or lower-adjacent grade to facilitate vertical linking investigation (i.e., the off-grade administration of items/tasks).

The response data for the items were collected from student samples ranging in size from 12,000 students in some high school grades to more than 40,000 in Grades 3 to 8. Though the samples were intended to be representative of their respective populations in characteristics such as their 2012 state test performance, gender, ethnicity, and special programs, the Pilot Test administration resulted in convenience student samples due to administration constraints. Because the representativeness of these samples is unknown, any comparisons based on results over grades and generalizations based on results of larger student populations should be regarded cautiously. Table 1 below summarizes the item pool sizes and student samples for all 18 tests.

Table 1. Summary of Number of Items and Students

| Grade | ELA | | Mathematics | |
| --- | --- | --- | --- | --- |
| | Number of Items | Number of Students | Number of Items | Number of Students |
| 3 | 241 | 41,450 | 212 | 41,502 |
| 4 | 236 | 49,797 | 214 | 43,722 |
| 5 | 184 | 49,522 | 210 | 46,406 |
| 6 | 227 | 49,670 | 213 | 42,051 |
| 7 | 210 | 44,430 | 230 | 41,408 |
| 8 | 232 | 41,132 | 224 | 44,650 |
| 9 | 146 | 25,690 | 135 | 19,298 |
| 10 | 157 | 16,079 | 139 | 12,438 |
| 11 | 287 | 18,904 | 306 | 24,405 |

After receipt of the scored student response data, statistical analyses of students' responses were conducted to gain information about the quality of the test questions. The analyses include several components: item difficulty, item discrimination, item response distribution, and differential item functioning (DIF). In general, items appeared difficult for the students who participated in the Pilot Test administration. Most items had average item score values below 0.5. There was a relatively small number of items that showed some performance differences between student groups. In addition to the item level statistics,

statistics for CAT item collections (modules) were computed, including the number of students taking the item collections, reliabilities, and observed score distributions as percentages of the maximum possible scores of the item collections (see Table 2 below). The median module score as a percentage of the module's maximum score shows that the items, when appearing as a collection, were on average difficult for Pilot administration participants. In general, the on-grade administration of Pilot CAT modules received more student responses than the off-grade administration of those test contents.

Table 2. Summary Statistics for CAT Item Collections (Modules)

| Subject | Grade | Student Samples | | Reliability | | | Percent of Maximum | | |
|---------|-------|------|------|------|------|--------|------|------|--------|
| | | Min | Max | Min | Max | Median | Min | Max | Median |
| ELA | 3 | 1,369 | 9,539 | 0.75 | 0.86 | 0.81 | 34.0 | 54.8 | 45.6 |
| | 4 | 1,092 | 7,426 | 0.70 | 0.83 | 0.77 | 34.8 | 54.4 | 44.8 |
| | 5 | 1,177 | 9,976 | 0.64 | 0.80 | 0.72 | 37.0 | 53.3 | 45.3 |
| | 6 | 1,278 | 4,915 | 0.60 | 0.80 | 0.72 | 37.3 | 48.3 | 43.0 |
| | 7 | 1,060 | 4,534 | 0.55 | 0.84 | 0.72 | 34.2 | 50.1 | 41.3 |
| | 8 | 491 | 4,331 | 0.53 | 0.79 | 0.69 | 35.1 | 46.4 | 42.4 |
| | 9 | 1,139 | 4,858 | 0.50 | 0.84 | 0.70 | 33.4 | 50.7 | 42.4 |
| | 10 | 507 | 2,838 | 0.64 | 0.81 | 0.72 | 31.4 | 47.1 | 36.8 |
| | 11 | 249 | 1,772 | 0.59 | 0.83 | 0.74 | 27.1 | 42.4 | 33.2 |
| Math | 3 | 1,743 | 6,199 | 0.67 | 0.87 | 0.79 | 26.0 | 51.8 | 36.8 |
| | 4 | 1,917 | 4,763 | 0.67 | 0.87 | 0.81 | 15.8 | 48.4 | 36.0 |
| | 5 | 2,062 | 5,116 | 0.74 | 0.86 | 0.83 | 23.7 | 42.5 | 35.6 |
| | 6 | 1,801 | 4,498 | 0.65 | 0.88 | 0.79 | 22.1 | 45.0 | 32.5 |
| | 7 | 893 | 3,642 | 0.62 | 0.84 | 0.79 | 15.6 | 36.0 | 26.1 |
| | 8 | 1,416 | 5,166 | 0.59 | 0.84 | 0.75 | 11.6 | 34.4 | 25.0 |
| | 9 | 705 | 3,527 | 0.58 | 0.76 | 0.63 | 9.9 | 26.6 | 20.9 |
| | 10 | 631 | 2,106 | 0.54 | 0.79 | 0.69 | 14.7 | 33.2 | 20.9 |
| | 11 | 536 | 2,272 | 0.52 | 0.83 | 0.72 | 10.0 | 28.3 | 18.9 |

Prior to conducting the dimensionality study and IRT analyses, the items were reviewed by content experts in light of these statistics. After the data review, more than 75% of ELA items and more than 83% of Math items were deemed appropriate for inclusion in dimensionality study and IRT analyses (except in Grade 9, where fewer than 70% of ELA items and fewer than 75% of Math items were included). Using the best available information from the Pilot, the evidence suggests that the unidimensionality model is consistently the preferred model. Therefore, the traditional IRT calibrations and linking can be performed. No changes are warranted to the scaling design, and all items for a grade and content area can be calibrated together simultaneously. Although a unidimensional model is consistently preferred, differences in dimensionality are most evident in Mathematics in the transition from Grade 8 to 9. This difference is somewhat expected since this delimits the transition into the course-specific content characterized by high school.

To support the IRT model selection process, analysis results are presented for IRT calibration evaluation, fit comparison, guessing evaluation, common discrimination evaluation, and ability estimates evaluation. Prior to conducting these analyses and in addition to establishing some item exclusion rules, a noteworthy

observation is that score categories for some items have to be pre-treated, because there are fewer than 10 examinees in some score categories. Of all the items that have received category collapsing due to sparse responses (Tables B.1 and B.2), more than 70% of them have fewer than 1,500 valid responses from the Pilot Test administration. Because the Field Test, like the Pilot Test, will use newly developed items, it is advisable, in order to mitigate the cases of score category collapsing, that the item-level sample sizes be larger for the IRT models that Smarter Balanced will adopt for Field Test analyses.

The model comparison analysis results with Pilot Test data suggest that the 2PL/GPC model combination should be adopted as the IRT model combination for calibrating Smarter Balanced items and establishing vertical scales. The 2PL/GPC model provides flexibility for estimating a range of item discriminations, without the complications of implementing a 3PL/GPC model. The major limitation of the 2PL/GPC model in this setting is that it has not been previously used for vertical scaling in K–12 assessments. This recommendation should be evaluated with caution given the experimental nature and limitations of the Pilot data, the possible change of item formats from Pilot to Field Test to operational administration, and the lack of information about vertical scaling results for the three models.

# Table of Contents

## Background

The Pilot Test administration was designed to collect data on items and tasks before they are used operationally and to make adjustments to test and item specifications based on the data collected. The Pilot Test was also intended to familiarize states, schools, teachers, and students with the kinds of items and tasks that will be a part of the Smarter Balanced summative assessments. All Pilot Test contents (CAT[1] items and PT) were administered via computer to a sample of students from across the member states of the consortium. The student responses were scored by automated computer algorithm, artificial intelligence scoring, or human raters. Multiple raters scored the same student response when the items were designed to be human scored.

Students participating in the Pilot Test took either a Mathematics test or an ELA test, including both the CAT and the PT components. The combination of a CAT component and a PT component covered the full content standards for the Pilot Test blueprints.

The CAT component consists of single-selection Selected-Response (SR) items, multiple-selection Selected-Response (MSR) items, and Constructed-Response (CR) items that will eventually contribute to the operational Summative and Interim CAT item pools. During the Pilot, the CAT component items were administered as linear tests via computer. CAT items were arranged into collections called modules, which were the basis for all analyses. Each CAT module of Pilot items contained a small number of items that were necessary to cover the content standards, which ranged from 23 to 33 items. Each participating student was administered at least one on-grade CAT module. Items were targeted at a given grade level for on-grade calibration. Items were also given to adjacent, off-grade students for vertical scaling.

A performance task (PT) is a collection of related items belonging to a common theme that consists of multiple items/observations and corresponding scores. Scores on the PT items ranged from 0 to 4. Most students were administered individually based performance tasks. Some were administered classroom-based performance tasks that contain some provision for classroom collaboration. An individually based performance task required that students approach the task independently without preparatory activities. A classroom-based performance task entails classroom activities or student interactions concerning a shared task. Although small-group work may be involved in some part of a task, it will not be scored, and preparatory activities were standardized to the maximum extent possible. All Pilot performance tasks were developed with a detachable classroom activity which means a PT could be administered with or without the classroom activity portion. Each item configuration was treated as a unique item for purposes of analysis. There was not enough information and data to compare the properties of the classroom and individual versions of the same performance tasks.

---

[1]Note that CAT refers to linear (fixed-form) administrations of the items in the item pool that will eventually be used for the computerized adaptive administration.

There are three Claims in the case of Math and four in ELA Note that in the CAT and performance task sample the Claims appear in different proportions, with some Claims overlapping. Tables 3 and 4 show the content assignments for ELA and Mathematics and the associated reporting categories.

Table 3. ELA Content Structure

| Component | Claims | Score Reporting Category |
|---|---|---|
| CAT Module | Reading | Literary |
| | | Informational |
| | Writing | Purpose/Focus/Organization |
| | | Evidence |
| | | Conventions |
| | Speaking/Listening | Listen/Interpret |
| | Research | Research |
| PT | Writing | Purpose/Focus/Organization |
| | | Evidence |
| | | Conventions |
| | Research | Research |

Table 4. Math Content Structure

| Component | Claims | Score Reporting Category |
|---|---|---|
| CAT Module | Concepts and Procedures | Domain 1 |
| | | Domain 2 |
| | Problem Solving | Problem Solving |
| | Modeling and Data Analysis | Modeling Data |
| | Communicating/Reasoning | Communicating/Reasoning |
| PT | Problem Solving/Modeling and Data Analysis | Problem Solving |
| | | Modeling Data |
| | Communicating/Reasoning | Communicating/Reasoning |

For the purpose of vertical scaling, the CAT component and PT component assigned to a student in a given grade could be on-grade or off-grade from either the adjacent-lower or upper grade. See Figure 1 for a depiction of content assignment by grade. The off-grade content was determined by content experts to be grade-level appropriate and representative of the construct; this also minimized opportunity-to-learn concerns to the maximum extent possible.

| Examinees by Grade | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Test Content by Grade | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 3 | 3 | 3 | | | | | | | |
| 4 | 4 | 4 | 4 | | | | | | |
| 5 | | 5 | 5 | 5 | | | | | |
| 6 | | | 6 | 6 | 6 | | | | |
| 7 | | | | 7 | 7 | 7 | | | |
| 8 | | | | | 8 | 8 | 8 | | |
| 9 | | | | | | 9 | 9 | 9 | 9 |
| 10 | | | | | | | 10 | 10 | 10 |
| 11 | | | | | | | | 11 | 11 |

Figure 1. Test Content Designation by Grade

The various data collection designs were proposed for the Pilot testing in 2013 (see *Pilot Test and Vertical Design,* Educational Testing Service, 2013). In accordance with the testing times that were established for the CAT and PT components and because of the need to control the number of items each student will take, the following Pilot Test data collection designs were adopted for Pilot Test administration.

- ELA tests will use the combination of alternate design variation 4 (Figure 2) and the supplemental design (Figure 4). Student samples selected for taking the ELA Pilot Tests will take either two CAT modules or one CAT module and one randomly-spiralled performance task.
- Mathematics tests will use the combination of alternate design variation 3 (Figure 3) and the supplemental design (Figure 4). Student samples selected for taking the math tests will take either three CAT modules or one CAT module and one randomly-spiralled performance task.

Figure 2. Alternate Data Collection Design Variation 4 (Adopted for ELA)

Figure 3. Alternate Data Collection Design Variation 3 (Adopted for Mathematics)

Figure 4. Supplemental Pilot Design (Adopted for both ELA and Mathematics)

The Pilot Tests were administered to students from February to May 2013. Student responses were scored in phases to facilitate analyses and item data review. The first phase of scored data files to support CAT item analyses and data review became available in October 2013. Item and DIF analyses were completed in October to support item data review. The second phase of complete data files with PT scored responses became available in January 2014; these were used to conduct the IRT and dimensionality related analyses that are documented in the later sections of this document.

Caution should be exercised when interpreting the Pilot analysis results due to the following constraints observed in the data in addition to those general limitations mentioned in the Executive Summary.

- The students who were tested for the Pilot administration resulted in a convenience student sample of the consortium.
- Most items that were administered in the Pilot administration were deemed not suitable for operational administration.
- Not all student responses were scored. The scoring had a maximum limit of 1,800 responses per item.
- Not all responses by a student were scored which means a student could have answered 50 items but only 30 were scored. This is the combined effect of scoring 1,800 responses per item limit and scoring by item instead of complete student records.
- Some content that was designed to be administered off-grade was not administered.
- In some cases, there were no performance tasks in the vertical linking anchors because most performance tasks were determined to be not scorable.

The following sections contain the procedures and results related to the dimensionality study and IRT model comparison.

## 1. Dimensionality Study

Before undertaking the Pilot calibration and scaling, Smarter Balanced sought insight concerning test dimensionality that will affect the IRT scaling design and ultimately the composite score that denotes overall student proficiency. This section describes the procedures used and outcomes pertaining to the dimensionality study based on the Pilot Test administration.

Math and ELA are scaled using multidimensional IRT for Grades 3 to 11, both within grade and across (adjacent) grades. Due to the mixed format data for the Smarter Balanced assessments containing SR and CR items, both unidimensional and multidimensional versions of the 2PL (M-2PL) and 2PPC (Yen, 1993) (M-2PPC) IRT scaling models are used. Both unidimensional and multidimensional models are compared using a number of model fit measures.

### 1.1 Rationale and Approach

As a factor analytic approach, multidimensional IRT (MIRT) is used to examine the dimensional structure. Table 5 below shows that there are two components to the dimensionality to be evaluated. The first component pertains to assessing the degree to which essential unidimensionality is met within a single grade and content area. The second aspect concerns the degree of invariance in the construct across two adjacent grades. Both criteria can be met or violated. A multidimensional composite of scores can be identified, but it should be consistent across grades in order to best support unidimensional scoring (Reckase, Ackerman, & Carlson, 1988).

Table 5. Dimensionality Analysis in the Context of Vertical and Horizontal Scaling

| | Construct Consistent Across Grades | |
|---|---|---|
| | Violated | Satisfied |
| **Unidimensionality within grades** Violated | 0,0 | 0,1 |
| Satisfied | 1,0 | 1,1 |

The dimensionality of the Pilot Test data is studied using MIRT. This MIRT approach has a number of advantages. First, MIRT is very close to the more familiar unidimensional IRT scaling techniques. This approach can utilize familiar unidimensional models as a starting point for model comparison. The baseline model is the unidimensional case in which other candidate models can be compared. Second, from a practical perspective the sparse data matrix used for unidimensional scaling can be leveraged without the need to create other types of data structures (i.e., covariance matrices). In addition, further insight can be obtained with respect to the vertical scaling. Using exploratory approaches, the shift in the nature of the construct across levels can be inspected across adjacent grade levels. Factor analysis here is primarily confirmatory in nature. The primary focus is the Claim structure for ELA and Mathematics. Simple structure refers to loading on a specified factor in a confirmatory approach. Complex structure refers to freeing items to load on multiple factors using an exploratory approach. By using an exploratory approach, the dimensional structure can be evaluated

graphically using item vectors. Global fit comparison will be undertaken to arrive at a preferred model that will be used to determine the scaling approach and the resulting score reporting. Both the overall model test fit (e.g., Bayesian Information Criterion) and graphical depictions using item vectors can be utilized in evaluating the factor structure.

The final reporting scale will be based on the Field Test administration. The Field Test and future operational administration will better reflect student performance while schools are in the process of transitioning instruction to the Common Core State Standards. The best case would be to replicate these findings in operational administrations.

## 1.2 Proposed Factor Models

The analysis consisted of two phases. The first phase examined each grade and content area separately (i.e., dimensionality within grade). The second phase investigated the dimensionality of two adjacent grade levels that contained unique grade specific items and common "vertical" linking items. The first step is a within-grade scaling. The results of the within-grade analysis will be evaluated before proceeding on to the across grades vertical linking. The next step is to concurrently scale two adjacent-grade tests and examine the resulting structure where a unidimensional multi-group model is implemented (Bock & Zimowski, 1997). The adjacent-grade levels have vertical linking items in common across grade groups. The choice in a candidate model can be assessed using the Bayesian or Akaike Inference Criterion (BIC/AIC) measures of global fit. The following factor models were proposed:

1) **Unidimensional Models:** The baseline model for comparison is the unidimensional version. Since unidimensional models are more constrained versions of multidimensional ones, MIRT software can be used to estimate them as well. The unidimensional versions will be implemented with the same calibration software to afford a similar basis of comparison with other multidimensional models.
2) **Multidimensional Models**
   - *Exploratory Models (Complex Structure).* The exploratory models "let the data speak" by adopting a complex structure in which items are permitted to load freely on multiple factors. Consistent with the approach outlined for unidimensional models, the first phase will examine each grade and content area separately (within-grade configuration). The next step is to concurrently scale two adjacent-grade test levels and examine the resulting structure. Using a two-dimensional exploratory model, item vectors can be evaluated graphically. An important aspect will be to note the direction of items and the composite vectors. If the same composite of factors is consistently present across grade levels, this will support the use of unidimensional approaches and the construction of the vertical scale.
   - *Confirmatory Models (Simple Structure).* Confirmatory models specify the loading of items on the factors, referred to as simple structure, according to specified criteria. Two types of confirmatory models will be investigated.
     A. *Claim Structure.* This model evaluates factors according to the Claim structure for each content area. For example, four Claims for Math are: Concepts & Procedures

(Domain 1 and Domain 2), Problem Solving, Modeling, and Communicating & Reasoning. A four factor model also results in ELA: Reading, Writing, Speaking/Listening, and Research.

B. *Bifactor Model.* A bifactor model is proposed in which an overall factor is proposed along with two or more minor ones. The minor factors will correspond to the Claim structure at each grade. A depiction of the bifactor model is given in Figure 5 consisting of a major factor and minor ones.
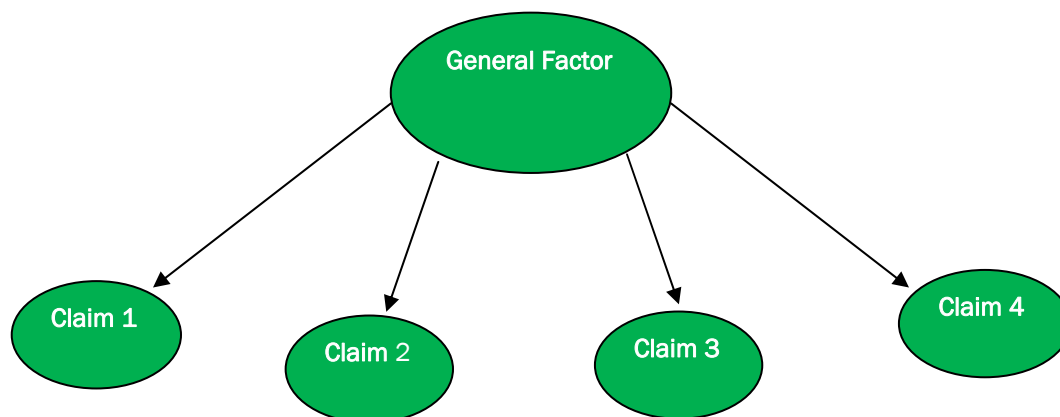


Figure 5. An Example of the Bifactor Model with Four Minor Factors Corresponding to Claims.

In total, four different models were evaluated for a grade and content area both within and across grades. The model and analysis configuration is summarized in Table 6 for the within-grade analysis and the across-grade configurations.

Table 6. Summary MIRT Analysis Configuration

| Model | Configuration | Content Areas | Grades | Total |
|---|---|---|---|---|
| Unidimensional | | | | |
| | Within grade | 2 | 9 | 18 |
| | Across grades | 2 | 8 | 16 |
| Multidimensional | | | | |
| Exploratory | Within grade | 2 | 9 | 18 |
| | Across grades | 2 | 8 | 16 |
| Claim Structure | Within grade | 2 | 9 | 18 |
| | Across grades | 2 | 8 | 16 |
| Bifactor | Within grade | 2 | 9 | 18 |
| | Across grades | 2 | 8 | 16 |
| Total | | | | 170 |

## 1.3 MIRT Scaling Models

With mixed data present in the Pilot Test, different types of IRT scaling models must be chosen. For SR items, the two-parameter logistic (2PL) model will be used or the M-2PL (McKinley & Reckase, 1983a) in the case of the multidimensional version. For CR items that includes all polytomous data, the two-parameter partial-credit model (2PPC) will be used. Likewise, for the dimensionality analysis the multidimensional two-parameter partial-credit model (M-2PPC) will be used (Yao & Schwarz, 2006). The multidimensional models used are compensatory in nature since high values for one theta (factor) can balance or help compensate for low ones in computing the probability of a response to an item for a student. The MIRT models chosen for the dimensionality analysis correspond to unidimensional models used for horizontal and vertical scaling of the Pilot Test. The M-2PL model for selected response is given below.

$$P_{ij} = 1 - \frac{1}{1 + e^{\vec{\beta}_{2j} \odot \vec{\theta}_i + \beta_{\delta j}}} = \frac{1}{1 + e^{-\vec{\beta}_{2j} \odot \vec{\theta}_i - \beta_{\delta j}}},$$

where $\vec{\beta}_{2j} = (\vec{\beta}_{2j1} \ldots \vec{\beta}_{2jD})$ is a vector of dimension D corresponding to items discrimination parameters, $\beta_{\delta j}$ is a scale difficult parameter and $\vec{\beta}_{2j} \odot \vec{\theta}_i = \sum_{l=1}^{D} \beta_{2jl} \theta_{il}$. For a polytomously scored item $j$, the probability of a response k-1 for an examinee with ability $\vec{\theta}_i$ is given by the multidimensional version of the 2-PPC model (Yao & Schwarz, 2006)

$$P_{ijk} = P(x_{ij} = k - 1 | \vec{\theta}_i \vec{\beta}_j) = \frac{e^{(k-1)\vec{\beta}_{2j} \odot \vec{\theta}_i^T - \sum_{t=1}^{k} \beta_{\delta tj}}}{\sum_{m=1}^{K_j} e^{((m-1)\vec{\beta}_{2j} \odot \vec{\theta}_i^T - \sum_{t=1}^{m} \beta_{\delta tj})}},$$

where $x_{ij=0,\ldots,K_j-1}$ is the response of examinee $i$ to item $j$, $\beta_{\delta jk}$ for k = 1,2...$K_j$ are threshold parameters, $\beta_{\delta 1j} = 0$ and $K_j$ is the number of response categories for the $j^{th}$ item.

## 1.4 Software and System Requirements

A scaling approach is needed that can implement models associated with mixed item types and one that makes provisions for missing data "not presented" by design. This "not-presented or not-reached" option is necessary since any student only takes a very small subset of the total available items. To be practical, the factor analysis needed to use the same data structures used for the traditional unidimensional IRT modeling. A wide variety of scaling models are implemented by **BMIRT** necessary for scaling mixed item types. The program also produces model fit and multigroup (i.e., across-grade) analysis. The **BMIRT** program (Yao, 2003) implements a full Bayesian approach to parameter estimation that uses the Metropolis-Hastings algorithm. Using a batch file approach to implement the program makes **BMIRT** efficient to implement across many grades and content areas. The **R** package (Weeks, 2010) **plink** performs multidimensional linking and other types of functions such as plotting of item characteristic curves. Other supporting **R** programs have been developed that check the Markov Chain Monte Carlo (MCMC) stationary process.

For parameter estimation, 1,000 MCMC iterations were used with 250 discarded for the MCMC burn-in. The resulting item parameters were then used as start values for another 1,000 MCMC cycles; 250 were discarded from these iterations as well. These second sets of iterations were used to compute the final parameter estimates. Note that 0.4 was used for the covariance for the prior ability functions (abilityPriorCov). Values of 0.0 corresponding to no relationship between factors and 0.8 indicating high correlations between factors were also evaluated. The difference in fit using these two other values was very small compared with the covariance of 0.4. **BMIRT** program defaults were used for other priors or proposal functions.

### 1.5 Evaluation of the Number and Types of Dimensions and MIRT Item Statistics

A primary method for evaluating models is to use overall test fit indices. The Bayesian and Akaike Information Criterion (Akaike, 1973; Schwarz, 1978) provided by **BMIRT** is given below.

$$BIC_k = G_k^2 + 2\log(N)\, df_k$$

$$AIC_k = G_k^2 + 2df_k$$

where $G_k^2$ is the likelihood and $2\log(N)\, df_k$ and $2df_k$ are penalties imposed for adding extra parameters to the model. These fit statistics can be used to compare either nested or non-nested models. Lower values of AIC and BIC indicate a better fitting candidate model. A referred factor structure results when it demonstrates the minimum fit value among several competing models. This permits comparison of model fit between unidimensional and multidimensional versions. Despite considerable advances in the estimation of a variety of complex models, no clear criteria exist for model acceptance. Several criteria will be evaluated to determine if the expected inferences are supported. This process of model choice is somewhat judgmental. To warrant the expense and operational complications involved in implementing a multidimensional scaling model, the preponderance of information would need to demonstrate the data are strongly multidimensional and that this multidimensionality varies over grades.

In Tables 7 and 8, AIC, BIC, the likelihood, and degrees of freedom (df) are presented for ELA and Mathematics. These tables show the overall fit by grade configuration (within-grade). They show the fit measures for the unidimensional, exploratory, Claims scores and bifactor models. The second set of global fit measures in Tables 9 and 10 show the across (adjacent) grade analysis where data from two adjacent grades are used. The measures for overall fit (across adjacent grades) are given and for each grade separately. Based on AIC and BIC, the unidimensional model is consistently the preferred model. Somewhat surprisingly, the bifactor model did not improve on the fit given by the Claims model.

For example, using Grade 3 ELA, the value of AIC for the unidimensional model was 1,580,927, which is lower than the values for the exploratory, Claims scores, and bifactor models. The values for BIC are larger by definition and follow the same pattern as AIC with the unidimensional model as the preferred candidate model. For the across-grade fit that contained vertical linking items, the

unidimensional model was also substantiated. The comparative fit across-grade models followed the same pattern as the within-grade analysis.

Table 7. Fit Measures for ELA within Grade

| Grade | Model | AIC | BIC | Likelihood | df |
|---|---|---|---|---|---|
| 3 | Unidimensional | 1,580,927 | 1,941,833 | -748,655 | 41,809 |
| | Exploratory | 1,637,492 | 2,355,936 | -735,518 | 83,228 |
| | Claim Scores | 1,736,151 | 3,169,699 | -702,006 | 166,069 |
| | Bifactor | 1,847,184 | 3,638,297 | -716,101 | 207,491 |
| 4 | Unidimensional | 1,671,889 | 2,113,952 | -785,799 | 50,145 |
| | Exploratory | 1,743,604 | 2,624,177 | -771,915 | 99,887 |
| | Claim Scores | 1,874,179 | 3,631,798 | -737,715 | 199,374 |
| | Bifactor | 2,003,755 | 4,199,911 | -752,758 | 249,119 |
| 5 | Unidimensional | 1,269,024 | 1,707,627 | -584,728 | 49,784 |
| | Exploratory | 1,338,209 | 2,212,463 | -569,872 | 99,233 |
| | Claim Scores | 1,471,467 | 3,217,049 | -537,599 | 198,134 |
| | Bifactor | 1,600,465 | 3,781,726 | -552,647 | 247,586 |
| 6 | Unidimensional | 1,422,993 | 1,863,403 | -661,524 | 49,972 |
| | Exploratory | 1,500,371 | 2,378,010 | -650,603 | 99,583 |
| | Claim Scores | 1,639,063 | 3,391,185 | -620,724 | 198,808 |
| | Bifactor | 1,763,784 | 3,953,161 | -633,470 | 248,422 |
| 7 | Unidimensional | 1,310,456 | 1,699,799 | -610,484 | 44,744 |
| | Exploratory | 1,372,121 | 2,147,449 | -596,958 | 89,102 |
| | Claim Scores | 1,488,947 | 3,036,270 | -566,652 | 177,821 |
| | Bifactor | 1,605,914 | 3,539,248 | -580,775 | 222,182 |
| 8 | Unidimensional | 1,282,613 | 1,640,099 | -599,857 | 41,450 |
| | Exploratory | 1,344,545 | 2,056,117 | -589,766 | 82,506 |
| | Claim Scores | 1,457,239 | 2,877,012 | -563,999 | 164,621 |
| | Bifactor | 1,561,028 | 3,334,914 | -574,834 | 205,680 |
| 9 | Unidimensional | 723,096 | 934,581 | -335,611 | 25,937 |
| | Exploratory | 760,617 | 1,181,126 | -328,737 | 51,572 |
| | Claim Scores | 835,337 | 1,673,916 | -314,823 | 102,845 |
| | Bifactor | 898,965 | 1,946,592 | -320,999 | 128,483 |
| 10 | Unidimensional | 486,630 | 612,020 | -226,999 | 16,316 |
| | Exploratory | 511,248 | 759,553 | -223,314 | 32,310 |
| | Claim Scores | 552,276 | 1,046,435 | -211,837 | 64,301 |
| | Bifactor | 597,408 | 1,214,505 | -218,406 | 80,298 |
| 11 | Unidimensional | 724,846 | 877,585 | -342,958 | 19,465 |
| | Exploratory | 745,309 | 1,045,794 | -334,360 | 38,294 |
| | Claim Scores | 795,682 | 1,391,686 | -321,886 | 75,955 |
| | Bifactor | 837,513 | 1,581,289 | -323,969 | 94,787 |

Table 8. Fit Measures for Mathematics within Grade

| Grade | Model | AIC | BIC | Likelihood | df |
|-------|-------|-----|-----|------------|-----|
| 3 | Unidimensional | 1,243,707 | 1,595,231 | -581,019 | 40,835 |
|   | Exploratory | 1,293,666 | 1,993,571 | -565,528 | 81,305 |
|   | Claim Scores | 1,415,106 | 2,811,801 | -545,305 | 162,248 |
|   | Bifactor | 1,521,203 | 3,266,305 | -557,881 | 202,721 |
| 4 | Unidimensional | 1,361,780 | 1,744,943 | -636,775 | 44,115 |
|   | Exploratory | 1,420,052 | 2,182,896 | -622,197 | 87,829 |
|   | Claim Scores | 1,560,890 | 3,083,122 | -605,185 | 175,260 |
|   | Bifactor | 1,671,350 | 3,573,288 | -616,698 | 218,977 |
| 5 | Unidimensional | 1,614,121 | 2,023,211 | -760,281 | 46,780 |
|   | Exploratory | 1,664,992 | 2,479,752 | -739,327 | 93,169 |
|   | Claim Scores | 1,818,934 | 3,445,061 | -723,517 | 185,950 |
|   | Bifactor | 1,919,462 | 3,951,285 | -727,389 | 232,342 |
| 6 | Unidimensional | 1,245,624 | 1,612,386 | -580,395 | 42,417 |
|   | Exploratory | 1,301,437 | 2,031,746 | -566,257 | 84,462 |
|   | Claim Scores | 1,444,817 | 2,902,243 | -553,853 | 168,555 |
|   | Bifactor | 1,540,013 | 3,361,011 | -559,403 | 210,603 |
| 7 | Unidimensional | 1,123,242 | 1,476,898 | -520,561 | 41,060 |
|   | Exploratory | 1,186,090 | 1,889,973 | -511,323 | 81,722 |
|   | Claim Scores | 1,318,147 | 2,722,512 | -496,025 | 163,049 |
|   | Bifactor | 1,419,308 | 3,173,927 | -505,940 | 203,714 |
| 8 | Unidimensional | 1,182,794 | 1,574,880 | -546,363 | 45,034 |
|   | Exploratory | 1,243,004 | 2,023,755 | -531,827 | 89,675 |
|   | Claim Scores | 1,398,606 | 2,956,713 | -520,343 | 178,960 |
|   | Bifactor | 1,496,807 | 3,443,605 | -524,800 | 223,604 |
| 9 | Unidimensional | 516,180 | 670,072 | -238,530 | 19,560 |
|   | Exploratory | 536,809 | 842,454 | -229,557 | 38,848 |
|   | Claim Scores | 612,138 | 1,221,311 | -228,642 | 77,427 |
|   | Bifactor | 648,848 | 1,409,797 | -227,706 | 96,718 |
| 10 | Unidimensional | 367,643 | 462,355 | -171,071 | 12,750 |
|   | Exploratory | 382,795 | 569,806 | -166,223 | 25,175 |
|   | Claim Scores | 425,940 | 797,570 | -162,942 | 50,028 |
|   | Bifactor | 454,729 | 918,679 | -164,909 | 62,456 |
| 11 | Unidimensional | 505,284 | 703,857 | -228,087 | 24,555 |
|   | Exploratory | 543,836 | 936,293 | -223,388 | 48,530 |
|   | Claim Scores | 630,439 | 1,410,687 | -218,736 | 96,483 |
|   | Bifactor | 683,748 | 1,657,903 | -221,413 | 120,461 |

Table 9. Fit Measures for ELA across Adjacent Grades

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|---|---|---|---|---|---|---|
| 3 to 4 | Unidimensional | Overall | 3,255,135 | 4,123,262 | -1,535,423 | 92,145 |
| | Unidimensional | 3 | 1,582,366 | 1,944,195 | -749,267 | 41,916 |
| | Unidimensional | 4 | 1,672,770 | 2,115,573 | -786,156 | 50,229 |
| | Exploratory | Overall | 3,381,393 | 5,108,951 | -1,507,330 | 183,367 |
| | Exploratory | 3 | 1,637,806 | 2,357,468 | -735,534 | 83,369 |
| | Exploratory | 4 | 1,743,587 | 2,625,139 | -771,796 | 99,998 |
| | Claim Scores | Overall | 3,703,214 | 7,149,559 | -1,485,804 | 365,803 |
| | Claim Scores | 3 | 1,734,620 | 3,169,972 | -701,032 | 166,278 |
| | Claim Scores | 4 | 1,968,594 | 3,727,544 | -784,772 | 199,525 |
| | Bifactor | Overall | 4,057,828 | 8,363,605 | -1,571,889 | 457,025 |
| | Bifactor | 3 | 1,850,234 | 3,643,444 | -717,383 | 207,734 |
| | Bifactor | 4 | 2,207,595 | 4,405,267 | -854,506 | 249,291 |
| 4 to 5 | Unidimensional | Overall | 2,942,383 | 3,894,243 | -1,371,059 | 100,132 |
| | Unidimensional | 4 | 1,672,823 | 2,115,732 | -786,170 | 50,241 |
| | Unidimensional | 5 | 1,269,560 | 1,709,105 | -584,889 | 49,891 |
| | Exploratory | Overall | 3,084,751 | 4,980,134 | -1,342,989 | 199,387 |
| | Exploratory | 4 | 1,742,772 | 2,624,456 | -771,373 | 100,013 |
| | Exploratory | 5 | 1,341,979 | 2,217,475 | -571,616 | 99,374 |
| | Claim Scores | Overall | 3,446,338 | 7,228,691 | -1,325,280 | 397,889 |
| | Claim Scores | 4 | 1,870,656 | 3,629,915 | -735,768 | 199,560 |
| | Claim Scores | 5 | 1,575,682 | 3,322,982 | -589,512 | 198,329 |
| | Bifactor | Overall | 3,837,936 | 8,563,813 | -1,421,824 | 497,144 |
| | Bifactor | 4 | 2,004,632 | 4,202,692 | -752,981 | 249,335 |
| | Bifactor | 5 | 1,833,305 | 4,016,530 | -668,843 | 247,809 |
| 5 to 6 | Unidimensional | Overall | 2,693,333 | 3,643,487 | -1,246,701 | 99,966 |
| | Unidimensional | 5 | 1,269,703 | 1,709,283 | -584,956 | 49,895 |
| | Unidimensional | 6 | 1,423,631 | 1,864,913 | -661,744 | 50,071 |
| | Exploratory | Overall | 2,842,088 | 4,734,451 | -1,221,948 | 199,096 |
| | Exploratory | 5 | 1,342,161 | 2,217,736 | -571,698 | 99,383 |
| | Exploratory | 6 | 1,499,927 | 2,378,712 | -650,251 | 99,713 |
| | Claim Scores | Overall | 3,202,642 | 6,979,344 | -1,203,973 | 397,348 |
| | Claim Scores | 5 | 1,468,207 | 3,215,798 | -535,741 | 198,362 |
| | Claim Scores | 6 | 1,734,435 | 3,488,126 | -668,231 | 198,986 |
| | Bifactor | Overall | 3,594,141 | 8,313,051 | -1,300,592 | 496,478 |
| | Bifactor | 5 | 1,603,426 | 3,787,039 | -553,860 | 247,853 |
| | Bifactor | 6 | 1,990,715 | 4,181,881 | -746,732 | 248,625 |

Table 9. Fit Measures for ELA across Adjacent Grades, continued

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|---|---|---|---|---|---|---|
| 6 to 7 | Unidimensional | Overall | 2,734,953 | 3,632,171 | -1,272,554 | 94,923 |
| | Unidimensional | 6 | 1,423,768 | 1,865,024 | -661,816 | 50,068 |
| | Unidimensional | 7 | 1,311,185 | 1,701,494 | -610,737 | 44,855 |
| | Exploratory | Overall | 2,869,962 | 4,656,033 | -1,246,020 | 188,961 |
| | Exploratory | 6 | 1,498,621 | 2,377,370 | -649,602 | 99,709 |
| | Exploratory | 7 | 1,371,341 | 2,147,975 | -596,419 | 89,252 |
| | Claim Scores | Overall | 3,228,796 | 6,792,497 | -1,237,369 | 377,029 |
| | Claim Scores | 6 | 1,635,272 | 3,389,033 | -618,642 | 198,994 |
| | Claim Scores | 7 | 1,593,524 | 3,142,710 | -618,727 | 178,035 |
| | Bifactor | Overall | 3,580,506 | 8,033,060 | -1,319,186 | 471,067 |
| | Bifactor | 6 | 1,766,238 | 3,957,518 | -634,481 | 248,638 |
| | Bifactor | 7 | 1,814,268 | 3,749,752 | -684,705 | 222,429 |
| 7 to 8 | Unidimensional | Overall | 2,595,184 | 3,403,519 | -1,211,203 | 86,389 |
| | Unidimensional | 7 | 1,311,172 | 1,701,351 | -610,746 | 44,840 |
| | Unidimensional | 8 | 1,284,012 | 1,642,351 | -600,457 | 41,549 |
| | Exploratory | Overall | 2,712,594 | 4,320,731 | -1,184,431 | 171,866 |
| | Exploratory | 7 | 1,368,710 | 2,145,152 | -595,125 | 89,230 |
| | Exploratory | 8 | 1,343,883 | 2,056,577 | -589,306 | 82,636 |
| | Claim Scores | Overall | 3,037,992 | 6,245,658 | -1,176,184 | 342,812 |
| | Claim Scores | 7 | 1,488,031 | 3,037,025 | -566,002 | 178,013 |
| | Claim Scores | 8 | 1,549,961 | 2,971,268 | -610,181 | 164,799 |
| | Bifactor | Overall | 3,357,227 | 7,364,695 | -1,250,324 | 428,289 |
| | Bifactor | 7 | 1,608,923 | 3,544,206 | -582,055 | 222,406 |
| | Bifactor | 8 | 1,748,304 | 3,523,941 | -668,269 | 205,883 |
| 8 to 9 | Unidimensional | Overall | 2,007,224 | 2,623,188 | -935,996 | 67,616 |
| | Unidimensional | 8 | 1,283,475 | 1,642,125 | -600,153 | 41,585 |
| | Unidimensional | 9 | 723,748 | 936,000 | -335,843 | 26,031 |
| | Exploratory | Overall | 2,106,595 | 3,330,799 | -918,914 | 134,384 |
| | Exploratory | 8 | 1,346,541 | 2,059,674 | -590,583 | 82,687 |
| | Exploratory | 9 | 760,054 | 1,181,582 | -328,330 | 51,697 |
| | Claim Scores | Overall | 2,355,982 | 4,796,591 | -910,079 | 267,912 |
| | Claim Scores | 8 | 1,454,408 | 2,876,536 | -562,310 | 164,894 |
| | Claim Scores | 9 | 901,573 | 1,741,563 | -347,769 | 103,018 |
| | Bifactor | Overall | 2,592,106 | 5,640,955 | -961,373 | 334,680 |
| | Bifactor | 8 | 1,564,631 | 3,341,268 | -576,317 | 205,999 |
| | Bifactor | 9 | 1,027,475 | 2,076,717 | -385,057 | 128,681 |

Table 9. Fit Measures for ELA across Adjacent Grades, continued

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|---|---|---|---|---|---|---|
| 9 to 10 | Unidimensional | Overall | 1,211,766 | 1,578,699 | -563,413 | 42,470 |
| | Unidimensional | 9 | 723,694 | 935,849 | -335,828 | 26,019 |
| | Unidimensional | 10 | 488,071 | 614,499 | -227,585 | 16,451 |
| | Exploratory | Overall | 1,274,759 | 2,001,981 | -553,209 | 84,171 |
| | Exploratory | 9 | 761,797 | 1,183,186 | -329,218 | 51,680 |
| | Exploratory | 10 | 512,962 | 762,658 | -223,990 | 32,491 |
| | Claim Scores | Overall | 1,417,427 | 2,865,158 | -541,149 | 167,565 |
| | Claim Scores | 9 | 833,651 | 1,673,535 | -313,821 | 103,005 |
| | Claim Scores | 10 | 583,776 | 1,079,925 | -227,328 | 64,560 |
| | Bifactor | Overall | 1,561,259 | 3,369,279 | -571,364 | 209,266 |
| | Bifactor | 9 | 899,379 | 1,948,523 | -321,021 | 128,669 |
| | Bifactor | 10 | 661,880 | 1,281,275 | -250,343 | 80,597 |
| 10 to 11 | Unidimensional | Overall | 1,213,870 | 1,518,346 | -570,955 | 35,980 |
| | Unidimensional | 9 | 487,682 | 613,971 | -227,408 | 16,433 |
| | Unidimensional | 10 | 726,188 | 879,570 | -343,547 | 19,547 |
| | Exploratory | Overall | 1,261,019 | 1,860,730 | -559,642 | 70,868 |
| | Exploratory | 9 | 513,973 | 763,477 | -224,520 | 32,466 |
| | Exploratory | 10 | 747,047 | 1,048,380 | -335,121 | 38,402 |
| | Claim Scores | Overall | 1,375,980 | 2,566,093 | -547,354 | 140,636 |
| | Claim Scores | 9 | 552,391 | 1,048,348 | -211,660 | 64,535 |
| | Claim Scores | 10 | 823,589 | 1,420,740 | -335,694 | 76,101 |
| | Bifactor | Overall | 1,485,638 | 2,970,985 | -567,295 | 175,524 |
| | Bifactor | 9 | 598,001 | 1,217,196 | -218,430 | 80,571 |
| | Bifactor | 10 | 887,637 | 1,632,715 | -348,865 | 94,953 |

Table 10. Fit Measures for Mathematics across Adjacent Grades

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|--------|-------|-------|-----|-----|------------|-----|
| 3 to 4 | Unidimensional | Overall | 2,609,055 | 3,402,805 | -1,219,552 | 84,976 |
| | Unidimensional | 3 | 1,245,590 | 1,597,234 | -581,946 | 40,849 |
| | Unidimensional | 4 | 1,363,465 | 1,746,733 | -637,606 | 44,127 |
| | Exploratory | Overall | 2,724,905 | 4,305,109 | -1,193,282 | 169,171 |
| | Exploratory | 3 | 1,299,575 | 1,999,652 | -568,463 | 81,325 |
| | Exploratory | 4 | 1,425,330 | 2,188,322 | -624,819 | 87,846 |
| | Claim Scores | Overall | 3,024,199 | 6,177,237 | -1,174,546 | 337,553 |
| | Claim Scores | 3 | 1,417,002 | 2,813,971 | -546,221 | 162,280 |
| | Claim Scores | 4 | 1,607,197 | 3,129,542 | -628,326 | 175,273 |
| | Bifactor | Overall | 3,226,816 | 7,166,308 | -1,191,660 | 421,748 |
| | Bifactor | 3 | 1,521,641 | 3,267,069 | -558,061 | 202,759 |
| | Bifactor | 4 | 1,705,175 | 3,607,218 | -633,599 | 218,989 |
| 4 to 5 | Unidimensional | Overall | 2,981,009 | 3,836,472 | -1,399,584 | 90,921 |
| | Unidimensional | 4 | 1,364,880 | 1,748,122 | -638,316 | 44,124 |
| | Unidimensional | 5 | 1,616,129 | 2,025,368 | -761,268 | 46,797 |
| | Exploratory | Overall | 3,086,050 | 4,789,382 | -1,361,990 | 181,035 |
| | Exploratory | 4 | 1,427,225 | 2,190,182 | -625,770 | 87,842 |
| | Exploratory | 5 | 1,658,825 | 2,473,795 | -736,220 | 93,193 |
| | Claim Scores | Overall | 3,436,284 | 6,835,279 | -1,356,887 | 361,255 |
| | Claim Scores | 4 | 1,564,470 | 3,086,883 | -606,954 | 175,281 |
| | Claim Scores | 5 | 1,871,814 | 3,498,151 | -749,933 | 185,974 |
| | Bifactor | Overall | 3,637,368 | 7,884,233 | -1,367,315 | 451,369 |
| | Bifactor | 4 | 1,673,654 | 3,575,809 | -617,825 | 219,002 |
| | Bifactor | 5 | 1,963,715 | 3,995,757 | -749,490 | 232,367 |
| 5 to 6 | Unidimensional | Overall | 2,867,813 | 3,705,554 | -1,344,691 | 89,215 |
| | Unidimensional | 5 | 1,617,910 | 2,027,052 | -762,169 | 46,786 |
| | Unidimensional | 6 | 1,249,902 | 1,616,768 | -582,522 | 42,429 |
| | Exploratory | Overall | 2,975,243 | 4,643,466 | -1,309,964 | 177,657 |
| | Exploratory | 5 | 1,669,399 | 2,484,237 | -741,521 | 93,178 |
| | Exploratory | 6 | 1,305,844 | 2,036,300 | -568,443 | 84,479 |
| | Claim Scores | Overall | 3,309,818 | 6,638,931 | -1,300,376 | 354,533 |
| | Claim Scores | 5 | 1,823,344 | 3,449,602 | -725,707 | 185,965 |
| | Claim Scores | 6 | 1,486,474 | 2,944,012 | -574,669 | 168,568 |
| | Bifactor | Overall | 3,497,384 | 7,656,979 | -1,305,717 | 442,975 |
| | Bifactor | 5 | 1,920,776 | 3,952,756 | -728,028 | 232,360 |
| | Bifactor | 6 | 1,576,608 | 3,397,710 | -577,689 | 210,615 |

Table 10. Fit Measures for Mathematics across Adjacent Grades, continued

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|--------|-------|-------|-----|-----|------------|-----|
| 6 to 7 | Unidimensional | Overall | 2,373,141 | 3,151,522 | -1,103,081 | 83,489 |
|  | Unidimensional | 6 | 1,247,380 | 1,614,177 | -581,269 | 42,421 |
|  | Unidimensional | 7 | 1,125,761 | 1,479,486 | -521,812 | 41,068 |
|  | Exploratory | Overall | 2,494,563 | 4,044,090 | -1,081,079 | 166,202 |
|  | Exploratory | 6 | 1,305,116 | 2,035,476 | -568,090 | 84,468 |
|  | Exploratory | 7 | 1,189,447 | 1,893,434 | -512,990 | 81,734 |
|  | Claim Scores | Overall | 2,803,345 | 5,895,090 | -1,070,052 | 331,620 |
|  | Claim Scores | 6 | 1,448,121 | 2,905,634 | -555,496 | 168,565 |
|  | Claim Scores | 7 | 1,355,223 | 2,759,640 | -514,557 | 163,055 |
|  | Bifactor | Overall | 2,985,167 | 6,848,058 | -1,078,251 | 414,333 |
|  | Bifactor | 6 | 1,546,176 | 3,367,277 | -562,473 | 210,615 |
|  | Bifactor | 7 | 1,438,991 | 3,193,645 | -515,778 | 203,718 |
| 7 to 8 | Unidimensional | Overall | 2,310,404 | 3,115,824 | -1,069,098 | 86,104 |
|  | Unidimensional | 7 | 1,125,531 | 1,479,238 | -521,699 | 41,066 |
|  | Unidimensional | 8 | 1,184,873 | 1,576,994 | -547,399 | 45,038 |
|  | Exploratory | Overall | 2,432,489 | 4,035,883 | -1,044,833 | 171,412 |
|  | Exploratory | 7 | 1,189,292 | 1,893,253 | -512,915 | 81,731 |
|  | Exploratory | 8 | 1,243,198 | 2,024,001 | -531,918 | 89,681 |
|  | Claim Scores | Overall | 2,758,373 | 5,957,640 | -1,037,167 | 342,020 |
|  | Claim Scores | 7 | 1,322,333 | 2,726,827 | -498,103 | 163,064 |
|  | Claim Scores | 8 | 1,436,040 | 2,994,112 | -539,064 | 178,956 |
|  | Bifactor | Overall | 2,946,770 | 6,944,012 | -1,046,057 | 427,328 |
|  | Bifactor | 7 | 1,424,973 | 3,179,747 | -508,754 | 203,732 |
|  | Bifactor | 8 | 1,521,798 | 3,468,526 | -537,303 | 223,596 |
| 8 to 9 | Unidimensional | Overall | 1,702,770 | 2,288,505 | -786,775 | 64,610 |
|  | Unidimensional | 8 | 1,184,158 | 1,576,296 | -547,039 | 45,040 |
|  | Unidimensional | 9 | 518,613 | 672,584 | -239,736 | 19,570 |
|  | Exploratory | Overall | 1,785,655 | 2,951,024 | -764,280 | 128,547 |
|  | Exploratory | 8 | 1,245,487 | 2,026,316 | -533,059 | 89,684 |
|  | Exploratory | 9 | 540,168 | 845,931 | -231,221 | 38,863 |
|  | Claim Scores | Overall | 2,027,808 | 4,352,371 | -757,491 | 256,413 |
|  | Claim Scores | 8 | 1,401,321 | 2,959,559 | -521,686 | 178,975 |
|  | Claim Scores | 9 | 626,486 | 1,235,746 | -235,805 | 77,438 |
|  | Bifactor | Overall | 2,160,865 | 5,065,062 | -760,082 | 320,350 |
|  | Bifactor | 8 | 1,504,595 | 3,451,549 | -528,675 | 223,622 |
|  | Bifactor | 9 | 656,270 | 1,417,297 | -231,407 | 96,728 |

Table 10. Fit Measures for Mathematics across Adjacent Grades, continued

| Grades | Model | Group | AIC | BIC | Likelihood | df |
|---|---|---|---|---|---|---|
| 9 to 10 | Unidimensional | Overall | 886,249 | 1,156,660 | -410,798 | 32,326 |
| | Unidimensional | 9 | 516,989 | 670,991 | -238,920 | 19,574 |
| | Unidimensional | 10 | 369,260 | 463,987 | -171,878 | 12,752 |
| | Exploratory | Overall | 922,509 | 1,458,271 | -397,207 | 64,047 |
| | Exploratory | 9 | 537,339 | 843,148 | -229,800 | 38,869 |
| | Exploratory | 10 | 385,170 | 572,203 | -167,407 | 25,178 |
| | Claim Scores | Overall | 1,052,414 | 2,118,811 | -398,726 | 127,481 |
| | Claim Scores | 9 | 614,092 | 1,223,540 | -229,584 | 77,462 |
| | Claim Scores | 10 | 438,322 | 809,885 | -169,142 | 50,019 |
| | Bifactor | Overall | 1,110,102 | 2,441,850 | -395,849 | 159,202 |
| | Bifactor | 9 | 649,857 | 1,411,136 | -228,168 | 96,760 |
| | Bifactor | 10 | 460,246 | 924,092 | -167,681 | 62,442 |
| 10 to 11 | Unidimensional | Overall | 876,674 | 1,194,819 | -400,926 | 37,411 |
| | Unidimensional | 10 | 369,223 | 464,151 | -171,832 | 12,779 |
| | Unidimensional | 11 | 507,452 | 706,648 | -229,094 | 24,632 |
| | Exploratory | Overall | 933,765 | 1,561,840 | -393,026 | 73,856 |
| | Exploratory | 10 | 388,458 | 575,789 | -169,011 | 25,218 |
| | Exploratory | 11 | 545,306 | 938,637 | -224,015 | 48,638 |
| | Claim Scores | Overall | 1,072,896 | 2,320,763 | -389,710 | 146,738 |
| | Claim Scores | 10 | 428,775 | 800,932 | -164,288 | 50,099 |
| | Claim Scores | 11 | 644,121 | 1,425,630 | -225,421 | 96,639 |
| | Bifactor | Overall | 1,141,252 | 2,699,050 | -387,443 | 183,183 |
| | Bifactor | 10 | 454,125 | 918,706 | -164,521 | 62,541 |
| | Bifactor | 11 | 687,127 | 1,662,746 | -222,922 | 120,642 |

## 1.6 MIRT Item Statistics and Graphs

Three primary MIRT item characteristics were computed that correspond to direction, difficulty, and discrimination presented graphically. The magnitude given by the length of the vector corresponds to its discriminating power

$$\sqrt{\vec{a}'\vec{a}}$$

The angle measure of the vector with each axis is

$$\alpha_{ij} = \arccos\frac{a_{ij}}{\sqrt{\vec{a}'\vec{a}}}$$

where $a_{ij}$ is the *j-th* element of the vector of item discriminations for item *i*. In order to obtain degrees, the angle measure is multiplied by 180/π. The quadrant of the plot in which an item resides roughly corresponds to its difficulty. The multidimensional difficulty is

$$\frac{-b_i}{\sqrt{\vec{a}'\vec{a}}},$$

where $b_i$ is the location or scalar item parameter related to item difficulty.

A composite directional vector can be computed using the matrix of discriminations **a** and then computing the eigenvalues for **a'a**. Each diagonal value in the matrix is the sum of the squared **a**-elements for each ability dimension of the matrix. The off diagonal values are the sums of the cross products of the **a**-elements from different dimensions. The eigenvector that corresponds to the largest eigenvalue is eigenvector one. The sum of the squared elements of the eigenvector is equal to one, and these elements have the properties of direction cosines. The direction cosines give the orientation of the reference composite with respect to the coordinate axes of the ability space. The angle between the reference composite and the coordinate axes can be determined by taking the arccosine of the elements of the eigenvector.

The Reckase, Martineau, & Kim (2000) item vector approach will be used to evaluate the characteristics of exploratory models using complex structure. The graphs showing the item vectors used the exploratory model with two dimensions. The development of these measures is conducted in a polar coordinate system so that direction can be specified as an angle from a particular axis. Using the MIRT item discrimination, the directions of maximum discrimination and MIRT item difficulty can all be depicted in the same graph. The origin of the item vectors is the MIRT difficulty. The reference composite vector composed of all items is also shown as a large red arrow. Item vectors that point in the same essential direction measure essentially the same dimension. Note that by definition, graphs of simple structure are not useful since all items are assigned to a defined axis corresponding to a factor.

The item vector plots are presented in Appendix A using the two-dimensional exploratory model. Plots are presented for ELA and Mathematics within grade, across two adjacent grades and for the subset of common, vertical linking items. The graphs of directional measures are presented in Figures[2] A.1 to A.9 for ELA. Figures A.11 to A.17 show item vectors for ELA across (adjacent) grades, and Figures A.18 to A.25 show them for the subset of vertical linking items. The graphs of directional measures are presented in Figures A.26 to A.34 for Mathematics. Figures A.35 to A.42 show item vectors for Mathematics across adjacent grades, and Figures A.43 to A.50 show them for the subset of linking items. The exploratory model is presented for diagnostic purposes to lend further insight into item functioning across dimensions. The plots for the exploratory model suggest that most items are primarily influenced by a composite of both factors. The item vector plot for Mathematics for the vertical linking items for Grades 8 and 9 shows the composite vector more closely associated with the first factor ($\theta_1$). This difference is reasonable since this delineates the transition to high school course specific content. In addition, for the vertical linking set for ELA Grades 9 and 10, some highly discriminating items are associated with the first factor.

## 1.7 Discussion and Conclusion

The evidence suggests that the unidimensional model was consistently the preferred model. This is consistent with the use of traditional IRT models for calibration and linking. No changes are warranted to the scaling design, and all items for a grade and content area can be calibrated together simultaneously. Although a unidimensional model was preferred, differences in dimensionality were most evident in Mathematics in the transition from Grade 8 to Grade 9. This difference is expected since this delimits the transition into the course specific content characterized by high school.

The approach adopted here is to use the best available information from the Pilot to inform decision-making regarding future development phases. At the minimum, the test dimensionality study based on the Pilot Test can only be viewed as preliminary and may need to be readdressed in the future. This is partly reflected in the changes that occurred in the item types, content configurations, and test design used in the Pilot compared with those employed for the Field Test. An overall concern is the degree of implementation of the Common Core State Standards across the Consortium. This will affect the results of this dimensionality study in ways that cannot currently be anticipated.

---

[2]The item vector plots represent separate calibrations.

## 2. Item Response Theory (IRT) Model Comparison

Within the family of IRT models there are two major choices to be made: 1) use of a unidimensional or multidimensional model, and 2) within the category of unidimensional models, the use of a Rasch one-parameter/partial credit model (Rasch/PC) combination, a two-parameter logistic/generalized partial credit model (2PL/GPC) combination, or a three-parameter logistic/generalized partial credit (3PL/GPC) combination.

It is highly desirable that a unidimensional model be used since the properties of these models are well known for scaling and are ones that have been used extensively in K–12 education to make critical decisions concerning students, teachers, and schools. Also, the IRT models selected must be implemented in operational Computerized Adaptive Testing (CAT). A multidimensional CAT with many constraints would be highly difficult to implement. Selection of an IRT model to comply with CAT implementation constraints may override other considerations.

The dimensionality study results from the previous section suggest that a unidimensional IRT model with a single IRT scale within each grade level could be used. Three unidimensional IRT model combinations are applied to the Pilot data for the dichotomous and polytomous item calibration. Specifically, these combinations are the Rasch one-parameter/partial credit model (1PL/PC) combination, the two-parameter logistic/generalized partial credit model (2PL/GPC) combination, and the three-parameter logistic/generalized partial credit model (3PL/GPC) combination. Calibration and scaling results based on all the three IRT model combinations are presented and compared, and they are used for making recommendations for IRT model choice for the Field Test and operational use and for determining the set of item parameter estimates to be stored in the item bank.

The Smarter Balanced assessment includes SR items, CR items, and performance task (PT) items that include both SR and CR items. For SR items, a 3PL, 2PL, or 1PL or Rasch model is used. The 3PL model is given by

$$P_i(\theta_j) = c_i + (1 - c_i) / (1 + \exp(-D a_i(\theta_j - b_i))),$$

where $P_i(\theta_j)$ is the probability of a correct response to item *i* by an examinee with ability $\theta_j$, and $a_i$, $b_i$, and $c_i$ are the discrimination, difficulty, and lower asymptote parameters, respectively, for item *i*, and *D* is a constant that puts the $\theta$ ability scale in the same metric as the normal ogive model (*D* = 1.7). The 3PL model can be constrained to equal the Rasch model by setting the constant *a* parameter to 1/*D* and the *c* parameter to 0. If the *a* parameter is left free to vary by item and c = 0, then the 2PL model results.

For CR items, the generalized partial credit model (Muraki, 1992) or partial credit model (Masters, 1982) is employed. The generalized partial credit model is given by

$$P_{ih}(\theta_j) = \frac{\exp(\sum_{v=1}^{h} Da_i(\theta_j - b_i + d_{iv}))}{\sum_{c=1}^{n_i} \exp(\sum_{v=1}^{c} Da_i(\theta_j - b_i + d_{iv}))} \quad ,$$

where $P_{ih}(\theta_j)$ is the probability of examinee *j* obtaining a score of *h* on item *i*, $n_i$ is the number of score categories item *i* contains, $b_i$ is the location parameter for item *i*, $d_{iv}$ is the category parameter for item *i* for category *v*, and *D* is a scaling constant. The generalized partial credit model can be constrained to equal the partial credit model by setting the *a* parameter to 1/*D*. The generalized partial credit model is equivalent to the two-parameter partial credit model used in the dimensionality study in the previous section (Yen and Fitzpatrick, 2006).

The choice of models within a unidimensional structure should take into account various considerations, including:

1. **Model simplicity or parsimony.** Model selection should balance goodness-of-fit and model simplicity. The Rasch model, which is easier to work with than the 2PL/GPC and 3PL/GPC, has worked well in many K–12 applications.
2. **Model fit.** Because the 3PL/GPC is a more general model, it provides better statistical model fit than the 2PL/GPC and the 1PL/PC; the 2PL/GPC provides better fit than 1PL/PC. (Often the improvement in fit from 2PL to 3PL can be far smaller than from 1PL to 2PL [Haberman, 2010]). However, statistical model fit, by itself, is not a sufficient basis for model choice. The practical implications of model choice should also be considered. For example, for CAT administration that aims to deliver items targeted at a specific student's ability level, fit of the IRT curve in the middle range may be more consequential than fit of the curve at the two ends. The primary practical implication of model misfit is a systematic difference between observed and predicted item characteristic functions, which affects the accuracy of scoring (i.e., the relationship of raw scores and trait estimates). Some item properties that affect model fit include:
    o Discriminations that vary systematically by item difficulty or trait level. The Rasch model assumes that item discrimination is uncorrelated with item difficulty. By examining plots or correlations of item discrimination versus item difficulty for the 2PL/GPC one can determine if the Rasch assumption is suitable for the Smarter tests. This result is also relevant to vertical scaling, since item discriminations for the same items administered across grade levels affect the vertical scaling.
    o Discriminations that vary systematically by item type (SR versus CR), number of score categories, or Claim areas. CR items with multiple score levels and/or CR scores based on the sum of multiple raters might be expected to have variant discrimination and not be adequately represented by the Rasch model (Sykes & Yen, 2000; Fitzpatrick, Link, Yen, Burket, Ito, & Sykes, 1996). Again, the results of the 2PL/GPC can be examined to see if there is a systematic relationship between item type/number of score categories/claim area and item discrimination.

3. **Model stability.** Findings from Holland (1990) indicated that unconstrained 3PL models must be expected to have stability problems. His study revealed that in the typical case of a standard normal prior, a unidimensional IRT model for dichotomous responses can be approximated by a log-linear model with only main effects and interactions. For a test of q items, the approximation is determined by 2q parameters, while 3PL model would require 3q parameters. This stability issue can be readily addressed by having appropriate priors on the c parameters, including holding them constant at logical values, particularly when sample sizes are small.

4. **Reasonableness of the vertical scale.** Since the selected IRT model will be used to establish a vertical scale, it is important to evaluate the reasonableness of the vertical scale, including expected growth from one grade to another, before making final decisions on the model for adoption. As suggested by research, the choice of the IRT scaling model may shrink or stretch out a measurement scale (Yen, 1981), and will subsequently impact the measurement of growth (Briggs & Weeks, 2009). Both the Rasch and 3PL have been used for developing K–12 vertical scales, and in the last two decades their scale properties have been broadly accepted by K–12 users (Yen & Fitzpatrick, 2006).

To support the Consortium in the IRT model selection process, the following results, including dimensionality analysis, IRT calibration, fit comparison, guessing evaluation, common discrimination evaluation and ability estimates evaluation results, are provided using the data collected in the Pilot administration. Both ELA and Math results are described. However, Math PT items are not included in the analysis. A considerable portion of the vertical linking items administered to upper grade levels show reverse growth patterns, which may be related to common core implementation progress. Given these vertical linking item issues, it is not possible to evaluate the reasonableness of the vertical scale as part of the model comparison analyses. For this reason, vertical scaling results are not provided as part of the model comparison analysis at this time.

## 2.1 Data Treatment

As indicated in the Background section of this memorandum, students took either multiple CAT components or a combination of CAT and PT components during the Pilot Test administration. The CAT or PT components administered might be on-grade or off-grade to facilitate vertical linking, but each participating student was administered at least one on-grade CAT module. PT items were included in the ELA IRT model comparison analyses but not in the Math analyses.

The first step was to create a sparse data matrix reflecting item scores as well as missing information by design. For a given grade, the dimension of the sparse matrix is the total number of students times the total number of unique items (i.e., scorable units). The remaining cells, representing items not administered to this student, have missing information indicated in the sparse matrix and are treated as "not presented" items in the IRT calibration. Data cleaning as described below was conducted before calibrations.

Data Cleaning Before Calibrations

The following item exclusion rules were followed:

a. Items that have no scored responses, or items that have scored responses in only one category were excluded;
b. CAT items that have on-grade item total correlations < 0.15 were removed from on-grade *AND* off-grade data sets regardless of their off-grade performance;
c. CAT items that have been recommended for "rejection" per content experts during data review meetings were removed from on-grade *AND* off-grade data sets;
d. PT items that have negative on-grade item total correlations were removed from on-grade *AND* off-grade data sets;
e. CAT or PT items with negative off-grade but reasonable on-grade item-total correlations were removed from the specific off-grade data sets only. For dimensionality studies, off-grade responses were calibrated together with on-grade responses.

The following category pre-treatments were followed:

a. Categories that have a reversed pattern of average criterion score progression (i.e., the average criterion score for a lower score category was higher than the average criterion score for a higher score category) at the on-grade level were collapsed in both on-grade *AND* off-grade data sets;
b. Categories with fewer than 10 examinees at on-grade level were collapsed with neighboring categories in both on-grade *AND* off-grade data sets. If the category that needed to be collapsed was a middle category, it was collapsed with the neighboring category that had lower student counts compared to the other neighboring category;
c. Categories that had a reversed pattern of average criterion score progression (i.e., the average criterion score for a lower score category was higher than the average criterion score for a higher score category) at the off-grade level but not at the on-grade level were collapsed in the specific off-grade data sets,
d. Categories with fewer than 10 examinees at the off-grade level but 10 or more examinees at the on-grade level were collapsed with neighboring categories in the specific off-grade data sets.

ELA and Math items that were dropped or received category pre-treatment before calibrations based on the above data cleaning procedure are listed in Tables B.1 and B.2 in Appendix B, respectively. These tables are presented separately in the appendix due to their lengths. Of all the items that required category collapsing due to sparse responses, more than 70% of them had fewer than 1,500 valid responses from the Pilot administration. This result emphasizes the importance of ensuring sufficient item-level sample size in the upcoming Field Test administration. The number of CAT/PT items that entered into ELA and Math IRT analyses after data cleaning and the examinee sample sizes associated with them are presented in Tables 11 and 12.

Table 11. Number of ELA Items in IRT Calibration and Examinee Sample Sizes at the Item Level

| Admin Grade | Item Grade | No. Items | | | Examinee Sample Size | |
|---|---|---|---|---|---|---|
| | | Total | CAT | PT | per CAT Item | per PT Item |
| 3 | 3 | 231 | 200 | 31 | 1,281–9,846 | 864–5,872 |
| | 4 | 48 | 44 | 4 | 1,377–6,641 | 1,408–1,413 |
| 4 | 3 | 48 | 44 | 4 | 1,101–2,501 | 1,293–1,304 |
| | 4 | 217 | 179 | 38 | 1,466–1,6343 | 897–3,518 |
| | 5 | 36 | 35 | 1 | 1,121–3,996 | 1,275 |
| 5 | 4 | 40 | 36 | 4 | 1,182–2,636 | 1,300–1,313 |
| | 5 | 175 | 144 | 31 | 1,420–18,373 | 950–2,797 |
| | 6 | 34 | 31 | 3 | 1,177–4,100 | 1,251–1,303 |
| 6 | 5 | 23 | 23 | | 1,278–4,048 | |
| | 6 | 202 | 161 | 41 | 1,399–12,760 | 929–3,577 |
| | 7 | 38 | 36 | 2 | 1,332–4,285 | 1,828–1,863 |
| 7 | 6 | 37 | 35 | 2 | 1,096–3,443 | 1,787–1,792 |
| | 7 | 195 | 163 | 32 | 1,378–12,078 | 1,066–3,835 |
| | 8 | 43 | 41 | 2 | 1,060–3,493 | 1,731–1,781 |
| 8 | 7 | 38 | 36 | 2 | 980–2,133 | 1,498–1,515 |
| | 8 | 202 | 168 | 34 | 1,084–13,077 | 1,074–3,867 |
| | 9 | 39 | 39 | | 492–1,076 | |
| 9 | 8 | 38 | 35 | 3 | 1,197–3,980 | 742–1,502 |
| | 9 | 126 | 80 | 46 | 4,583–5,008 | 553–720 |
| | 10 | 46 | 46 | | 1,139–1,374 | |
| 10 | 9 | 41 | 41 | | 507–615 | |
| | 10 | 133 | 109 | 24 | 1,382–3,013 | 369–527 |
| | 11 | 50 | 48 | 2 | 522–1,206 | 549–551 |
| 11 | 9 | 80 | 80 | | 256–322 | |
| | 10 | 107 | 107 | | 249–320 | |
| | 11 | 261 | 221 | 40 | 1,328–3,729 | 384–1,710 |

Table 12. Number of Math Items in IRT Calibration and Examinee Sample Sizes at the Item Level

| Admin Grade | Item Grade | No. CAT Items | Examinee Sample Size per CAT Item |
|---|---|---|---|
| 3 | 3 | 207 | 416–14,735 |
| | 4 | 47 | 1,743–3,582 |
| 4 | 3 | 38 | 1,917–4,373 |
| | 4 | 209 | 497–9,642 |
| | 5 | 37 | 1,941–4,355 |
| 5 | 4 | 41 | 2,129–4,636 |
| | 5 | 204 | 496–10,338 |
| | 6 | 39 | 2,062–4,607 |
| 6 | 5 | 41 | 1,838–4,030 |
| | 6 | 189 | 483–9,213 |
| | 7 | 48 | 1,807–3,939 |
| 7 | 6 | 41 | 912–1,992 |
| | 7 | 190 | 441–11,138 |
| | 8 | 37 | 952–2,148 |
| 8 | 7 | 33 | 1,422–3,292 |
| | 8 | 191 | 473–8,556 |
| | 9 | 47 | 1,416–3,280 |
| 9 | 8 | 23 | 1,273–2,775 |
| | 9 | 103 | 484–6497 |
| | 10 | 56 | 1,287–2,826 |
| 10 | 9 | 51 | 692–1,511 |
| | 10 | 122 | 493–3,889 |
| | 11 | 48 | 634–1,528 |
| 11 | 9 | 80 | 561–2,709 |
| | 10 | 90 | 536–2,910 |
| | 11 | 263 | 422–5,407 |

## 2.2 IRT Model Calibration

IRT calibration is conducted based on 1PL/PC, 2PL/GPC, and 3PL/GPC model combinations using PARSCALE (Muraki & Bock, 2003). PARSCALE properties are well known, and a variety of unidimensional IRT models can be implemented with it.

Item Treatment Rules During Calibration

Item treatment was conducted during calibration in situations of non-convergence or unreasonably large standard errors for item parameter estimates. Non-convergence was defined by either not achieving the criterion of largest parameter change lower than 0.005 or an erratic pattern of -2log likelihood values. Standard errors were evaluated against item parameter estimates as part of the reasonableness check procedure. Calibration issues in the Pilot analyses were found to be mostly caused by the following:

a. Local item dependence (LID). Some PT items with item ID ending with "a" and "b" (i.e., ID values such as "40583a" and "40583b") are highly correlated. These items involved the same student responses scored with different rubrics. The LID makes these items appear highly discriminating, thus causing problems for PARSCALE in locating the slope parameter estimates for these items.
b. Low item discrimination. While CAT items with item-total correlations lower than 0.15 have been removed from the pool, there are still some PT items with poor discrimination. Items with poor discrimination, especially ones that are difficult, sometimes cause convergence issues in 3PL calibration.
c. Guessing parameter indeterminacy in the 3PL. Guessing parameter starting values may cause issues in a 3PL calibration, sometimes leading to large standard errors for difficulty estimates (> 1.0) or unreasonable guessing parameter estimates (zero guessing parameter estimates associated with standard errors larger than 0.04).

To address these calibration issues and allow smooth estimation, the following item treatments were made at the individual item level, when a specific item was identified as being problematic.

*For SR Items:*

a. For the 3PL model, guessing parameter starting values were changed for the item. First the guessing parameter starting value was changed to .25, next to .10, and next to 0, if calibration issues persisted.
b. For the 3PL model, the guessing parameter was held at a fixed value if changing the guessing parameter starting value did not solve the calibration issues. The guessing parameter was first fixed to .25, next to .10, and next to 0 if calibration issues persisted.
c. If none of the above item treatments solved the calibration issue, then the item was removed.

*For CR Items:*

    a. Change starting values for the item. For polytomous items, there is an option to use category starting values that are constant values from "scores for ordinal or ranked data" instead of the PARSCALE default category starting values.

    b. Collapse categories for the polytomous item.

    c. If none of the above item treatments solved a calibration issue, then the item was removed. Usually when PARSCLE was having a convergence issue due to LID, one item out of the pair that caused LID was removed.

Items that received treatment during IRT calibration based on the above-described treatment steps are listed in Tables B.3 and B.4 in Appendix B for the ELA and Math tests, respectively. Note that no items were deleted from the 1PL analyses and a few items were deleted from the 2PL analyses, largely due to LID issues. Additional item treatment was made in 3PL analyses due to c-parameter estimation issues. Thus, there were some differences in the items included in the following results for the three models.

Under each model combination, the convergence process, IRT parameter estimates as well as standard errors associated with them, and item goodness-of-fit analyses results were used to evaluate the quality of the resulting item and ability parameter estimates. In general, convergence under each IRT model combination was reached and the resulting IRT item/ability parameter estimates under each model combination were reasonable.

## 2.3 IRT Model Comparisons

### Fit Comparison

To allow comparison of fit across different IRT model combinations, PARSCALE $G^2$ statistics were evaluated. In PARSCALE, a likelihood ratio $G^2$ test statistic can be used to compare the frequencies of correct and incorrect responses in the intervals on the $\theta$ continuum with those expected based on the fitted model (du Toit, 2003)

$$G_i^2 = 2\sum_{h=1}^{n_g}\left[ r_{ih} \log_e \frac{r_{ih}}{N_h P_i(\overline{\theta}_h)} + (N_h - r_{ih})\log_e \frac{N_h - r_{ih}}{N_h(1 - P_i(\overline{\theta}_h))}\right],$$

where $n_g$ is the total number of intervals, $r_{ih}$ is the observed frequency of correct responses to item *i* in interval *h*, $N_h$ is the number of examinees in interval *h*, $\overline{\theta}_h$ is the average ability of examinees in interval *h*, and $P_i(\overline{\theta}_h)$ is the value of the fitted response function for item *i* at $\overline{\theta}_h$. In PARSCALE, $G^2$ statistics are calculated and presented in Phase 2 output.

Since the $G^2$ statistic tends to be sensitive to sample size (i.e., flagging more items under larger sample sizes), it is used as a descriptive statistic in this study instead of a statistic for significance

testing. Since there are many items for any grade/content area combination, the distributions of $G^2$ are compared across IRT model combinations. Tables 13 and 14 present the summary of $G^2$ statistics across 1PL/PC, 2PL/GPC, and 3PL/GPC models for ELA and Math, respectively. Although $G^2$ statistics may not be strictly comparable across models due to the difference in degrees of freedom, the size of the $G^2$ statistics in general may still provide some evidence for comparing fit across models considering that the degree of freedom for each item is roughly comparable across different models. The tables show that for most of the tests the mean value of $G^2$ for the 1PL/PC is substantially greater than the mean values for the other two model combinations, indicating considerable average improvement in fit with 2PL/GPC and 3PL/GPC in comparison with 1PL/PC.

Table 13. Summary of $G^2$ Statistics of On-Grade ELA Items across 1PL, 2PL, and 3PL IRT Models

| Item Grade | 1PL/PC | | | 2PL/GPC | | | 3PL/GPC | | |
|---|---|---|---|---|---|---|---|---|---|
| | No. of Items | $G^2$ Mean | $G^2$ SD | No. of Items | $G^2$ Mean | $G^2$ SD | No. of Items | $G^2$ Mean | $G^2$ SD |
| 3 | 231 | 151 | 114 | 231 | 79 | 58 | 231 | 79 | 60 |
| 4 | 217 | 128 | 93 | 216 | 72 | 38 | 216 | 70 | 41 |
| 5 | 175 | 121 | 87 | 171 | 75 | 42 | 171 | 73 | 43 |
| 6 | 202 | 132 | 99 | 197 | 79 | 51 | 197 | 78 | 51 |
| 7 | 195 | 127 | 87 | 190 | 84 | 57 | 190 | 84 | 58 |
| 8 | 202 | 135 | 118 | 199 | 85 | 73 | 199 | 84 | 73 |
| 9 | 126 | 103 | 67 | 119 | 72 | 44 | 119 | 72 | 45 |
| 10 | 133 | 93 | 56 | 129 | 63 | 31 | 129 | 62 | 33 |
| 11 | 261 | 79 | 48 | 259 | 57 | 34 | 259 | 57 | 35 |

Table 14. Summary of $G^2$ Statistics of On-Grade Math Items across 1PL, 2PL, and 3PL IRT Models

| Item Grade | 1PL/PC | | | 2PL/GPC | | | 3PL/GPC | | |
|---|---|---|---|---|---|---|---|---|---|
| | No. of Items | $G^2$ Mean | $G^2$ SD | No. of Items | $G^2$ Mean | $G^2$ SD | No. of Items | $G^2$ Mean | $G^2$ SD |
| 3 | 207 | 127 | 88 | 207 | 86 | 58 | 207 | 84 | 58 |
| 4 | 209 | 139 | 99 | 209 | 92 | 82 | 209 | 90 | 84 |
| 5 | 204 | 167 | 127 | 204 | 95 | 77 | 204 | 93 | 80 |
| 6 | 189 | 145 | 106 | 189 | 96 | 69 | 189 | 93 | 69 |
| 7 | 190 | 162 | 123 | 190 | 113 | 94 | 190 | 110 | 97 |
| 8 | 191 | 152 | 111 | 191 | 110 | 86 | 191 | 114 | 99 |
| 9 | 103 | 111 | 66 | 103 | 95 | 62 | 103 | 94 | 60 |
| 10 | 122 | 97 | 52 | 122 | 71 | 42 | 122 | 71 | 44 |
| 11 | 263 | 72 | 58 | 263 | 72 | 88 | 263 | 68 | 74 |

## Guessing Evaluation

The single-selection SR items in the Pilot Test had four answer choices. Since 1PL and 2PL models assume minimal guessing, the amount of guessing involved for SR items is evaluated by examining the size of guessing parameter estimates under the 3PL/GPC model combinations. Large guessing parameter estimates provide evidence for the use of 3PL models and small guessing parameter estimates allow possible use of 1PL and 2PL models. Tables 15 and 16 present the mean, standard deviation, minimum, maximum, and range of guessing parameter estimates for items administered on-grade for ELA and Math, respectively. Results indicate that the average guessing is below .20 for most tests. The range of the guessing values showed a consistent pattern across grade levels in that the majority of SR items had guessing parameter estimates below .20 but greater than .10.

Table 15. Summary of Guessing Parameter Estimates for On-Grade ELA Items

| Item Grade | Admin Grade | No. of Items | $c$ Estimates Summary | | | | $c$ Estimates Range | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | SD | Min | Max | 0—0.10 | 0.10—0.20 | 0.20—0.30 | >0.30 |
| 3 | 3 | 76 | 0.16 | 0.07 | 0.06 | 0.39 | 16 | 43 | 14 | 3 |
| 4 | 4 | 111 | 0.17 | 0.07 | 0.04 | 0.36 | 20 | 53 | 31 | 7 |
| 5 | 5 | 77 | 0.15 | 0.07 | 0.00 | 0.31 | 16 | 40 | 20 | 1 |
| 6 | 6 | 75 | 0.15 | 0.07 | 0.05 | 0.33 | 23 | 35 | 14 | 3 |
| 7 | 7 | 76 | 0.18 | 0.07 | 0.06 | 0.38 | 9 | 39 | 25 | 3 |
| 8 | 8 | 77 | 0.15 | 0.07 | 0.00 | 0.34 | 16 | 46 | 10 | 5 |
| 9 | 9 | 36 | 0.16 | 0.08 | 0.04 | 0.31 | 10 | 15 | 9 | 2 |
| 10 | 10 | 46 | 0.16 | 0.08 | 0.00 | 0.35 | 9 | 24 | 10 | 3 |
| 11 | 11 | 91 | 0.18 | 0.07 | 0.04 | 0.39 | 12 | 48 | 25 | 6 |

Table 16. Summary of Guessing Parameter Estimates for On-Grade Math Items

| Item Grade | Admin Grade | No. of Items | $c$ Estimates Summary | | | | $c$ Estimates Range | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | SD | Min | Max | 0—0.10 | 0.10—0.20 | 0.20—0.30 | >0.30 |
| 3 | 3 | 34 | 0.18 | 0.07 | 0.05 | 0.36 | 3 | 21 | 8 | 2 |
| 4 | 4 | 31 | 0.17 | 0.06 | 0.03 | 0.29 | 3 | 18 | 10 | 0 |
| 5 | 5 | 39 | 0.18 | 0.10 | 0.02 | 0.43 | 13 | 10 | 11 | 5 |
| 6 | 6 | 41 | 0.21 | 0.09 | 0.08 | 0.38 | 5 | 14 | 13 | 9 |
| 7 | 7 | 31 | 0.20 | 0.08 | 0.07 | 0.39 | 3 | 12 | 13 | 3 |
| 8 | 8 | 34 | 0.18 | 0.07 | 0.07 | 0.32 | 3 | 18 | 10 | 3 |
| 9 | 9 | 14 | 0.20 | 0.08 | 0.09 | 0.35 | 1 | 8 | 3 | 2 |
| 10 | 10 | 19 | 0.26 | 0.11 | 0.06 | 0.46 | 2 | 3 | 8 | 6 |
| 11 | 11 | 32 | 0.19 | 0.08 | 0.05 | 0.37 | 4 | 15 | 9 | 4 |

## Common Discrimination Evaluation

The Rasch model assumes a common item discrimination across all items. Analyses were conducted to evaluate if item discrimination varied systematically by item difficulty, item type (SR vs. CR), number of item score categories, or item claim areas. This evaluation was done by plotting item discrimination versus item difficulty estimates from the 2PL/GPC model calibrations across all items within each grade level, with items of different types (i.e., SR vs CR, items with different numbers of score categories, items in different claim areas) highlighted. When the distribution of item discrimination is reasonably homogeneous, the selection of a model that assumes equal item discrimination may be viable.

Tables 17 and 18 summarize discrimination and difficulty parameter estimates and correlations between them under the 2PL/GPC for ELA and Math items administered on-grade. These summary statistics are provided for the overall set of items as well as groups of items characterized by item type (SR/CR), score categories (number of discrete possible score points), and claim areas. Figures B.1 and B.2[3] in Appendix B present, for ELA and Math and at each grade level, plots of item discrimination versus item difficulty under the 2PL/GPC with item type, score category, and claim area highlighted for each item. Results show that for the 2PL/GPC model there is moderate negative correlation between item difficulty and discrimination for ELA. There is less evidence for neither positive nor negative correlation between item difficulty and discrimination for Math items.

Tables 17 and 18 also show sizable standard deviations for discrimination parameter estimates, above 0.20 for all subjects and grade levels, which indicate a substantially wide range of discrimination parameter estimates for the items in the pool. The average discriminations vary somewhat, but not considerably, across item groupings. The CR items were slightly more discriminating on average than SR items. The pattern of item discrimination across different numbers of score categories was inconsistent across subjects. For ELA, items with 2 and 3 score categories had comparable discrimination, while items with 4 score categories generally had higher average discrimination (which might be due to local item dependence issues for PT items). For Math, the fewer the number of score categories, the higher the item discrimination. ELA items in Claim areas 2 and 4 had slightly higher average discriminations than items in claim areas 1 and 3 for most of the grade levels. Math items did not show a noticeable pattern of differential discrimination across different Claim areas.

An advantage of the 2PL/GPC in comparison to the 1PL/PC is that it would permit using items with a range of item discriminations, while the 1PL/PC could flag items with both very high and very low discriminations for exhibiting poor fit and requiring further content review.

---

[3]These plots have inconsistent ranges because the range may be unnecessarily wide for most of the tests as a result of the fewer tests having quite extreme difficulty and discrimination values. When the range is wide, the scatter points would all be in the middle and it will be difficult to identify any patterns if they exist.

Table 17. Summary of 2PL/GPC Slope and Difficulty Estimates and Correlations for On-Grade ELA Items

| Item Grade | Item Grouping | | No. of Items | *a* Estimates Summary | | | | *b* Estimates Summary | | | | *a* and *b* Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| 3 | Overall | | 231 | 0.63 | 0.23 | 0.15 | 1.24 | 0.32 | 1.22 | -1.87 | 5.00 | -0.29 |
| | Item Type | SR | 76 | 0.64 | 0.25 | 0.16 | 1.23 | -0.44 | 1.09 | -1.87 | 5.00 | -0.64 |
| | | CR | 155 | 0.62 | 0.22 | 0.15 | 1.24 | 0.69 | 1.11 | -1.80 | 4.35 | -0.12 |
| | Score Categories | 2 | 134 | 0.65 | 0.24 | 0.16 | 1.23 | 0.08 | 1.27 | -1.87 | 5.00 | -0.39 |
| | | 3 | 91 | 0.56 | 0.20 | 0.15 | 1.09 | 0.61 | 1.09 | -1.80 | 3.38 | -0.18 |
| | | 4 | 6 | 1.04 | 0.16 | 0.86 | 1.24 | 1.39 | 0.14 | 1.22 | 1.60 | -0.39 |
| | Claim Area | 1 | 85 | 0.63 | 0.23 | 0.18 | 1.12 | 0.10 | 1.12 | -1.84 | 3.14 | -0.51 |
| | | 2 | 64 | 0.68 | 0.26 | 0.18 | 1.24 | 0.33 | 0.99 | -1.25 | 2.98 | 0.03 |
| | | 3 | 44 | 0.60 | 0.21 | 0.15 | 1.06 | -0.22 | 0.96 | -1.87 | 2.23 | -0.24 |
| | | 4 | 38 | 0.57 | 0.22 | 0.16 | 1.06 | 1.44 | 1.40 | -1.80 | 5.00 | -0.41 |
| 4 | Overall | | 216 | 0.57 | 0.23 | 0.20 | 1.40 | 0.33 | 1.21 | -1.93 | 4.14 | -0.15 |
| | Item Type | SR | 111 | 0.54 | 0.21 | 0.20 | 1.24 | -0.32 | 0.89 | -1.93 | 2.18 | -0.59 |
| | | CR | 105 | 0.61 | 0.24 | 0.20 | 1.40 | 1.01 | 1.13 | -1.28 | 4.14 | -0.06 |
| | Score Categories | 2 | 148 | 0.56 | 0.21 | 0.20 | 1.24 | 0.00 | 1.12 | -1.93 | 3.54 | -0.30 |
| | | 3 | 59 | 0.53 | 0.21 | 0.20 | 1.26 | 0.97 | 1.16 | -1.28 | 4.14 | -0.11 |
| | | 4 | 9 | 1.02 | 0.25 | 0.73 | 1.40 | 1.48 | 0.44 | 1.01 | 2.00 | -0.91 |
| | Claim Area | 1 | 78 | 0.58 | 0.22 | 0.20 | 1.24 | -0.16 | 1.02 | -1.85 | 2.48 | -0.48 |
| | | 2 | 58 | 0.62 | 0.25 | 0.27 | 1.40 | 0.42 | 1.07 | -1.93 | 2.71 | 0.04 |
| | | 3 | 40 | 0.49 | 0.17 | 0.22 | 0.83 | -0.05 | 0.91 | -1.55 | 2.54 | -0.21 |
| | | 4 | 40 | 0.57 | 0.24 | 0.20 | 1.26 | 1.51 | 1.20 | -0.89 | 4.14 | -0.10 |
| 5 | Overall | | 171 | 0.61 | 0.20 | 0.19 | 1.15 | 0.34 | 1.21 | -2.14 | 3.38 | -0.16 |
| | Item Type | SR | 77 | 0.57 | 0.21 | 0.19 | 1.05 | -0.46 | 0.84 | -2.14 | 1.87 | -0.53 |
| | | CR | 94 | 0.63 | 0.18 | 0.20 | 1.15 | 1.00 | 1.06 | -1.06 | 3.38 | -0.16 |
| | Score Categories | 2 | 115 | 0.59 | 0.19 | 0.19 | 1.05 | -0.01 | 1.15 | -2.14 | 3.38 | -0.25 |
| | | 3 | 50 | 0.61 | 0.19 | 0.20 | 1.12 | 1.01 | 1.02 | -1.01 | 2.96 | -0.16 |
| | | 4 | 6 | 0.80 | 0.26 | 0.57 | 1.15 | 1.51 | 0.63 | 0.80 | 2.14 | -0.80 |
| | Claim Area | 1 | 55 | 0.56 | 0.18 | 0.19 | 0.92 | 0.21 | 1.15 | -1.98 | 2.90 | -0.20 |
| | | 2 | 51 | 0.62 | 0.20 | 0.27 | 1.15 | 0.39 | 1.05 | -1.74 | 2.75 | -0.07 |
| | | 3 | 32 | 0.62 | 0.20 | 0.28 | 1.05 | -0.51 | 0.84 | -2.14 | 1.42 | -0.59 |
| | | 4 | 33 | 0.65 | 0.21 | 0.20 | 1.12 | 1.31 | 1.18 | -1.13 | 3.38 | -0.18 |
| 6 | Overall | | 197 | 0.58 | 0.28 | 0.17 | 2.06 | 0.65 | 1.48 | -1.79 | 8.05 | -0.10 |
| | Item Type | SR | 75 | 0.51 | 0.20 | 0.17 | 1.01 | -0.31 | 0.98 | -1.79 | 2.65 | -0.54 |
| | | CR | 122 | 0.63 | 0.31 | 0.19 | 2.06 | 1.23 | 1.44 | -1.26 | 8.05 | -0.16 |
| | Score Categories | 2 | 128 | 0.58 | 0.25 | 0.17 | 1.34 | 0.41 | 1.47 | -1.79 | 5.29 | -0.11 |
| | | 3 | 66 | 0.58 | 0.29 | 0.19 | 2.06 | 1.06 | 1.44 | -1.26 | 8.05 | -0.15 |
| | | 4 | 3 | 1.09 | 0.61 | 0.59 | 1.77 | 1.59 | 0.24 | 1.40 | 1.86 | -0.46 |
| | Claim Area | 1 | 77 | 0.55 | 0.19 | 0.19 | 1.04 | 0.52 | 1.27 | -1.79 | 3.54 | -0.41 |
| | | 2 | 56 | 0.61 | 0.35 | 0.18 | 2.06 | 0.54 | 1.32 | -1.34 | 4.80 | -0.02 |

| Item Grade | Item Grouping | | No. of Items | a Estimates Summary | | | | b Estimates Summary | | | | a and b Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| | | 3 | 29 | 0.52 | 0.17 | 0.17 | 0.85 | -0.38 | 0.96 | -1.74 | 2.65 | -0.46 |
| | | 4 | 35 | 0.68 | 0.34 | 0.19 | 1.34 | 1.93 | 1.69 | -0.29 | 8.05 | -0.23 |
| 7 | Overall | | 190 | 0.53 | 0.21 | 0.11 | 1.18 | 0.57 | 1.34 | -2.25 | 6.61 | -0.30 |
| | Item Type | SR | 76 | 0.52 | 0.24 | 0.19 | 1.18 | -0.13 | 1.10 | -2.25 | 3.29 | -0.56 |
| | | CR | 114 | 0.53 | 0.19 | 0.11 | 1.14 | 1.04 | 1.29 | -1.76 | 6.61 | -0.18 |
| | Score Categories | 2 | 115 | 0.55 | 0.22 | 0.19 | 1.18 | 0.26 | 1.38 | -2.25 | 5.81 | -0.32 |
| | | 3 | 70 | 0.49 | 0.19 | 0.11 | 1.07 | 1.00 | 1.15 | -1.32 | 6.61 | -0.23 |
| | | 4 | 5 | 0.61 | 0.08 | 0.51 | 0.72 | 1.68 | 0.38 | 1.24 | 2.10 | 0.26 |
| | Claim Area | 1 | 70 | 0.52 | 0.20 | 0.12 | 1.18 | 0.47 | 1.41 | -2.21 | 5.81 | -0.38 |
| | | 2 | 46 | 0.52 | 0.16 | 0.21 | 0.96 | 0.40 | 1.23 | -2.25 | 2.71 | -0.14 |
| | | 3 | 42 | 0.47 | 0.23 | 0.19 | 1.06 | 0.29 | 1.13 | -1.90 | 3.29 | -0.60 |
| | | 4 | 32 | 0.61 | 0.23 | 0.11 | 1.14 | 1.42 | 1.31 | -0.25 | 6.61 | -0.38 |
| 8 | Overall | | 199 | 0.56 | 0.27 | 0.08 | 1.58 | 0.53 | 1.21 | -2.87 | 6.17 | -0.12 |
| | Item Type | SR | 77 | 0.50 | 0.20 | 0.08 | 1.02 | -0.11 | 0.98 | -2.01 | 2.53 | -0.50 |
| | | CR | 122 | 0.59 | 0.30 | 0.13 | 1.58 | 0.93 | 1.17 | -2.87 | 6.17 | -0.11 |
| | Score Categories | 2 | 119 | 0.56 | 0.24 | 0.08 | 1.26 | 0.23 | 1.17 | -2.01 | 6.17 | -0.17 |
| | | 3 | 74 | 0.49 | 0.24 | 0.13 | 1.25 | 0.93 | 1.17 | -2.87 | 4.47 | -0.19 |
| | | 4 | 6 | 1.24 | 0.35 | 0.69 | 1.58 | 1.49 | 0.21 | 1.30 | 1.83 | -0.26 |
| | Claim Area | 1 | 75 | 0.49 | 0.17 | 0.13 | 0.90 | 0.38 | 1.40 | -2.01 | 6.17 | -0.36 |
| | | 2 | 50 | 0.64 | 0.35 | 0.18 | 1.58 | 0.36 | 1.02 | -2.87 | 2.18 | 0.19 |
| | | 3 | 40 | 0.47 | 0.21 | 0.17 | 1.02 | 0.44 | 1.16 | -1.78 | 2.95 | -0.61 |
| | | 4 | 34 | 0.69 | 0.30 | 0.08 | 1.26 | 1.21 | 0.82 | -0.53 | 3.30 | -0.29 |
| 9 | Overall | | 119 | 0.60 | 0.24 | 0.20 | 1.20 | 0.64 | 1.33 | -2.24 | 6.04 | 0.01 |
| | Item Type | SR | 36 | 0.54 | 0.20 | 0.22 | 0.99 | -0.43 | 0.78 | -1.60 | 1.21 | -0.51 |
| | | CR | 83 | 0.63 | 0.25 | 0.20 | 1.20 | 1.10 | 1.25 | -2.24 | 6.04 | -0.03 |
| | Score Categories | 2 | 64 | 0.58 | 0.23 | 0.20 | 1.08 | 0.38 | 1.36 | -1.60 | 3.54 | 0.01 |
| | | 3 | 51 | 0.60 | 0.26 | 0.21 | 1.20 | 0.89 | 1.28 | -2.24 | 6.04 | -0.06 |
| | | 4 | 4 | 0.87 | 0.15 | 0.73 | 1.06 | 1.45 | 0.22 | 1.25 | 1.74 | -0.44 |
| | Claim Area | 1 | 56 | 0.58 | 0.27 | 0.20 | 1.20 | 0.54 | 1.46 | -2.24 | 6.04 | -0.13 |
| | | 2 | 25 | 0.65 | 0.20 | 0.29 | 1.00 | 0.46 | 1.11 | -1.60 | 2.61 | 0.07 |
| | | 3 | 14 | 0.49 | 0.19 | 0.28 | 0.99 | -0.19 | 1.08 | -1.23 | 2.12 | -0.46 |
| | | 4 | 24 | 0.67 | 0.23 | 0.22 | 1.10 | 1.55 | 0.86 | -0.30 | 3.30 | 0.24 |
| 10 | Overall | | 129 | 0.60 | 0.25 | 0.19 | 1.33 | 0.75 | 1.26 | -1.78 | 4.70 | -0.18 |
| | Item Type | SR | 46 | 0.56 | 0.25 | 0.22 | 1.11 | -0.10 | 0.92 | -1.78 | 2.78 | -0.55 |
| | | CR | 83 | 0.63 | 0.24 | 0.19 | 1.33 | 1.23 | 1.17 | -0.76 | 4.70 | -0.16 |
| | Score Categories | 2 | 73 | 0.61 | 0.24 | 0.22 | 1.12 | 0.53 | 1.40 | -1.78 | 4.70 | -0.21 |
| | | 3 | 53 | 0.57 | 0.24 | 0.19 | 1.32 | 1.02 | 1.01 | -0.76 | 3.25 | -0.17 |
| | | 4 | 3 | 1.00 | 0.30 | 0.73 | 1.33 | 1.53 | 0.22 | 1.28 | 1.71 | -0.99 |
| | Claim Area | 1 | 59 | 0.55 | 0.20 | 0.19 | 1.05 | 0.74 | 1.40 | -1.78 | 4.70 | -0.28 |

| Item Grade | Item Grouping | | No. of Items | a Estimates Summary | | | | b Estimates Summary | | | | a and b Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| | | 2 | 30 | 0.73 | 0.28 | 0.22 | 1.33 | 0.90 | 1.18 | -1.34 | 3.92 | -0.15 |
| | | 3 | 20 | 0.52 | 0.25 | 0.21 | 1.11 | 0.00 | 0.84 | -1.21 | 1.91 | -0.72 |
| | | 4 | 20 | 0.65 | 0.26 | 0.23 | 1.30 | 1.32 | 0.96 | -0.16 | 2.78 | -0.14 |
| | Overall | | 259 | 0.54 | 0.22 | 0.18 | 1.32 | 1.01 | 1.20 | -1.97 | 5.09 | -0.15 |
| | Item Type | SR | 91 | 0.49 | 0.17 | 0.19 | 0.91 | 0.24 | 0.89 | -1.97 | 2.85 | -0.55 |
| | | CR | 168 | 0.57 | 0.23 | 0.18 | 1.32 | 1.43 | 1.14 | -1.29 | 5.09 | -0.18 |
| | Score Categories | 2 | 142 | 0.54 | 0.20 | 0.19 | 1.21 | 0.75 | 1.26 | -1.97 | 5.09 | -0.09 |
| 11 | | 3 | 110 | 0.52 | 0.22 | 0.18 | 1.18 | 1.34 | 1.07 | -0.68 | 4.71 | -0.27 |
| | | 4 | 7 | 0.89 | 0.28 | 0.69 | 1.32 | 1.36 | 0.10 | 1.26 | 1.50 | -0.27 |
| | Claim Area | 1 | 95 | 0.47 | 0.20 | 0.19 | 1.19 | 1.20 | 1.26 | -1.97 | 4.83 | -0.22 |
| | | 2 | 54 | 0.65 | 0.24 | 0.26 | 1.32 | 0.65 | 1.00 | -1.25 | 2.92 | 0.04 |
| | | 3 | 65 | 0.48 | 0.16 | 0.18 | 0.86 | 0.72 | 1.17 | -1.29 | 4.71 | -0.46 |
| | | 4 | 45 | 0.65 | 0.20 | 0.26 | 1.21 | 1.49 | 1.13 | -0.52 | 5.09 | -0.02 |

Table 18. Summary of 2PL/GPC Slope and Difficulty Estimates and Correlations for On-Grade Math Items

| Item Grade | Item Grouping | | No. of Items | a Estimates Summary | | | | b Estimates Summary | | | | a and b Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| | Overall | | 207 | 0.69 | 0.21 | 0.21 | 1.31 | 0.31 | 1.43 | -4.06 | 4.42 | 0.01 |
| | Item Type | SR | 34 | 0.65 | 0.23 | 0.21 | 1.21 | -0.81 | 1.25 | -4.06 | 1.84 | -0.33 |
| | | CR | 173 | 0.69 | 0.21 | 0.21 | 1.31 | 0.53 | 1.36 | -3.16 | 4.42 | 0.04 |
| | Score Categories | 2 | 126 | 0.75 | 0.21 | 0.21 | 1.31 | 0.18 | 1.50 | -4.06 | 3.63 | 0.05 |
| 3 | | 3 | 66 | 0.61 | 0.16 | 0.32 | 0.98 | 0.45 | 1.28 | -2.77 | 4.42 | 0.08 |
| | | 4 | 15 | 0.50 | 0.18 | 0.21 | 0.77 | 0.77 | 1.36 | -1.68 | 3.25 | 0.28 |
| | Claim Area | 1 | 154 | 0.71 | 0.21 | 0.21 | 1.31 | 0.20 | 1.43 | -4.06 | 3.63 | 0.04 |
| | | 2 | 28 | 0.65 | 0.18 | 0.37 | 1.22 | 0.50 | 1.15 | -1.26 | 3.53 | 0.02 |
| | | 3 | 17 | 0.60 | 0.19 | 0.21 | 0.86 | 1.00 | 1.79 | -2.77 | 4.42 | -0.16 |
| | | 4 | 8 | 0.60 | 0.21 | 0.28 | 0.95 | 0.37 | 1.08 | -1.68 | 1.55 | 0.71 |
| | Overall | | 209 | 0.72 | 0.25 | 0.19 | 1.32 | 0.72 | 1.20 | -3.42 | 3.97 | 0.01 |
| | Item Type | SR | 31 | 0.63 | 0.25 | 0.19 | 1.10 | -0.09 | 1.36 | -1.91 | 3.86 | -0.58 |
| | | CR | 178 | 0.73 | 0.25 | 0.27 | 1.32 | 0.86 | 1.12 | -3.42 | 3.97 | 0.08 |
| | Score Categories | 2 | 141 | 0.78 | 0.25 | 0.19 | 1.32 | 0.70 | 1.28 | -3.42 | 3.97 | -0.02 |
| 4 | | 3 | 55 | 0.59 | 0.17 | 0.28 | 1.09 | 0.64 | 1.03 | -1.66 | 2.45 | 0.30 |
| | | 4 | 13 | 0.50 | 0.14 | 0.28 | 0.77 | 1.32 | 0.82 | 0.05 | 2.46 | 0.10 |
| | Claim Area | 1 | 158 | 0.72 | 0.24 | 0.24 | 1.32 | 0.54 | 1.23 | -3.42 | 3.58 | 0.09 |
| | | 2 | 30 | 0.70 | 0.28 | 0.19 | 1.22 | 1.23 | 0.91 | -0.10 | 3.86 | -0.30 |
| | | 3 | 14 | 0.72 | 0.31 | 0.28 | 1.26 | 1.36 | 0.98 | 0.01 | 3.97 | -0.26 |
| | | 4 | 7 | 0.70 | 0.20 | 0.50 | 1.08 | 1.41 | 0.62 | 0.40 | 2.44 | 0.30 |
| 5 | Overall | | 204 | 0.71 | 0.26 | 0.23 | 1.38 | 0.55 | 1.10 | -3.34 | 4.17 | 0.17 |

| Item Grade | Item Grouping | | No. of Items | a Estimates Summary | | | | b Estimates Summary | | | | a and b Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| | Item Type | SR | 39 | 0.62 | 0.21 | 0.23 | 1.13 | -0.11 | 0.73 | -1.83 | 1.72 | 0.00 |
| | | CR | 165 | 0.73 | 0.27 | 0.27 | 1.38 | 0.70 | 1.12 | -3.34 | 4.17 | 0.14 |
| | Score Categories | 2 | 144 | 0.76 | 0.27 | 0.23 | 1.38 | 0.47 | 1.15 | -3.34 | 3.43 | 0.25 |
| | | 3 | 53 | 0.60 | 0.20 | 0.27 | 1.11 | 0.71 | 0.99 | -1.29 | 4.17 | 0.04 |
| | | 4 | 7 | 0.47 | 0.09 | 0.30 | 0.58 | 0.82 | 0.70 | -0.26 | 1.68 | 0.48 |
| | Claim Area | 1 | 156 | 0.71 | 0.25 | 0.23 | 1.31 | 0.43 | 1.12 | -3.34 | 4.17 | 0.19 |
| | | 2 | 26 | 0.76 | 0.28 | 0.38 | 1.38 | 1.04 | 0.99 | -0.70 | 3.43 | -0.01 |
| | | 3 | 15 | 0.56 | 0.24 | 0.30 | 1.09 | 0.69 | 0.85 | -1.29 | 1.72 | -0.15 |
| | | 4 | 7 | 0.77 | 0.33 | 0.34 | 1.13 | 1.00 | 1.03 | -0.26 | 2.33 | 0.68 |
| | Overall | | 189 | 0.70 | 0.27 | 0.19 | 1.58 | 0.95 | 1.19 | -1.77 | 4.09 | 0.01 |
| 6 | Item Type | SR | 41 | 0.55 | 0.21 | 0.19 | 1.10 | 0.21 | 1.17 | -1.77 | 2.98 | -0.55 |
| | | CR | 148 | 0.74 | 0.27 | 0.20 | 1.58 | 1.16 | 1.12 | -1.54 | 4.09 | 0.00 |
| | Score Categories | 2 | 133 | 0.75 | 0.28 | 0.19 | 1.58 | 0.94 | 1.28 | -1.77 | 4.09 | 0.05 |
| | | 3 | 49 | 0.61 | 0.18 | 0.20 | 0.99 | 1.02 | 0.99 | -0.78 | 3.69 | -0.23 |
| | | 4 | 7 | 0.43 | 0.12 | 0.32 | 0.64 | 0.74 | 0.80 | -0.36 | 2.07 | 0.03 |
| | Claim Area | 1 | 149 | 0.68 | 0.27 | 0.19 | 1.58 | 0.84 | 1.21 | -1.77 | 4.09 | -0.08 |
| | | 2 | 22 | 0.74 | 0.21 | 0.41 | 1.17 | 1.03 | 1.13 | -1.19 | 3.05 | 0.48 |
| | | 3 | 11 | 0.63 | 0.26 | 0.30 | 1.10 | 1.84 | 0.73 | 0.63 | 3.10 | -0.05 |
| | | 4 | 7 | 0.97 | 0.32 | 0.63 | 1.50 | 1.84 | 0.75 | 0.67 | 2.97 | 0.33 |
| | Overall | | 190 | 0.66 | 0.26 | 0.15 | 1.43 | 1.38 | 1.19 | -1.81 | 6.38 | -0.11 |
| 7 | Item Type | SR | 31 | 0.46 | 0.15 | 0.23 | 0.91 | 0.82 | 1.12 | -1.81 | 3.84 | -0.68 |
| | | CR | 159 | 0.70 | 0.26 | 0.15 | 1.43 | 1.49 | 1.18 | -1.02 | 6.38 | -0.16 |
| | Score Categories | 2 | 101 | 0.73 | 0.27 | 0.23 | 1.43 | 1.38 | 1.03 | -1.81 | 4.20 | 0.11 |
| | | 3 | 74 | 0.60 | 0.22 | 0.15 | 1.15 | 1.40 | 1.40 | -1.02 | 6.38 | -0.45 |
| | | 4 | 15 | 0.50 | 0.21 | 0.21 | 0.96 | 1.25 | 1.11 | -0.62 | 3.89 | 0.07 |
| | Claim Area | 1 | 148 | 0.67 | 0.26 | 0.15 | 1.43 | 1.41 | 1.17 | -1.81 | 6.38 | -0.09 |
| | | 2 | 20 | 0.74 | 0.22 | 0.27 | 1.17 | 0.90 | 0.83 | -0.87 | 2.59 | 0.22 |
| | | 3 | 17 | 0.54 | 0.26 | 0.21 | 1.06 | 1.65 | 1.65 | -0.92 | 5.46 | -0.33 |
| | | 4 | 5 | 0.70 | 0.16 | 0.51 | 0.96 | 1.44 | 1.02 | 0.29 | 2.56 | 0.14 |
| | Overall | | 191 | 0.65 | 0.27 | 0.13 | 1.47 | 1.25 | 1.17 | -1.49 | 5.12 | -0.08 |
| 8 | Item Type | SR | 34 | 0.48 | 0.17 | 0.20 | 0.76 | 0.79 | 1.09 | -0.99 | 4.54 | -0.60 |
| | | CR | 157 | 0.69 | 0.28 | 0.13 | 1.47 | 1.35 | 1.16 | -1.49 | 5.12 | -0.09 |
| | Score Categories | 2 | 121 | 0.70 | 0.30 | 0.18 | 1.47 | 1.35 | 1.22 | -1.20 | 5.12 | -0.13 |
| | | 3 | 62 | 0.57 | 0.20 | 0.13 | 1.07 | 1.02 | 1.06 | -1.49 | 4.95 | -0.02 |
| | | 4 | 8 | 0.50 | 0.16 | 0.28 | 0.82 | 1.40 | 0.96 | 0.12 | 3.04 | -0.62 |
| | Claim Area | 1 | 149 | 0.63 | 0.27 | 0.13 | 1.45 | 1.20 | 1.17 | -1.49 | 5.12 | -0.11 |
| | | 2 | 26 | 0.74 | 0.31 | 0.34 | 1.47 | 1.58 | 1.26 | -0.97 | 5.12 | -0.05 |
| | | 3 | 12 | 0.65 | 0.16 | 0.48 | 0.88 | 1.04 | 1.00 | -1.01 | 2.49 | -0.15 |
| | | 4 | 4 | 0.72 | 0.25 | 0.45 | 1.04 | 1.50 | 0.42 | 0.97 | 1.85 | -0.18 |

| Item Grade | Item Grouping | | No. of Items | a Estimates Summary | | | | b Estimates Summary | | | | a and b Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| 9 | Overall | | 103 | 0.60 | 0.27 | 0.15 | 1.42 | 1.92 | 1.27 | -0.62 | 7.34 | 0.00 |
| | Item Type | SR | 14 | 0.46 | 0.20 | 0.21 | 0.77 | 0.99 | 1.04 | -0.29 | 3.76 | -0.62 |
| | | CR | 89 | 0.62 | 0.28 | 0.15 | 1.42 | 2.07 | 1.25 | -0.62 | 7.34 | -0.01 |
| | Score Categories | 2 | 63 | 0.68 | 0.28 | 0.21 | 1.42 | 1.89 | 1.31 | -0.62 | 7.34 | 0.09 |
| | | 3 | 34 | 0.50 | 0.21 | 0.15 | 1.02 | 1.94 | 1.18 | 0.29 | 6.10 | -0.14 |
| | | 4 | 6 | 0.33 | 0.09 | 0.23 | 0.44 | 2.13 | 1.54 | -0.44 | 3.80 | -0.30 |
| | Claim Area | 1 | 84 | 0.62 | 0.28 | 0.15 | 1.42 | 1.93 | 1.31 | -0.62 | 7.34 | -0.02 |
| | | 2 | 11 | 0.47 | 0.22 | 0.20 | 0.77 | 1.77 | 1.36 | -0.44 | 4.21 | 0.18 |
| | | 3 | 6 | 0.51 | 0.18 | 0.24 | 0.69 | 2.14 | 0.84 | 1.02 | 3.24 | -0.11 |
| | | 4 | 2 | 0.52 | 0.08 | 0.47 | 0.58 | 1.71 | 0.30 | 1.50 | 1.93 | -1.00 |
| 10 | Overall | | 122 | 0.67 | 0.36 | 0.17 | 1.76 | 1.32 | 1.10 | -1.15 | 5.49 | 0.12 |
| | Item Type | SR | 19 | 0.48 | 0.22 | 0.18 | 1.12 | 0.88 | 1.48 | -0.71 | 5.49 | -0.35 |
| | | CR | 103 | 0.71 | 0.37 | 0.17 | 1.76 | 1.40 | 1.00 | -1.15 | 3.84 | 0.15 |
| | Score Categories | 2 | 68 | 0.81 | 0.40 | 0.18 | 1.76 | 1.40 | 1.15 | -0.71 | 5.49 | 0.06 |
| | | 3 | 42 | 0.53 | 0.19 | 0.17 | 0.91 | 1.22 | 1.00 | -1.15 | 3.67 | 0.19 |
| | | 4 | 12 | 0.37 | 0.17 | 0.17 | 0.75 | 1.18 | 1.16 | -0.36 | 3.23 | 0.22 |
| | Claim Area | 1 | 94 | 0.69 | 0.38 | 0.17 | 1.76 | 1.13 | 1.08 | -1.15 | 5.49 | 0.21 |
| | | 2 | 13 | 0.67 | 0.22 | 0.26 | 1.10 | 1.80 | 0.69 | -0.02 | 2.54 | 0.03 |
| | | 3 | 10 | 0.50 | 0.32 | 0.17 | 1.33 | 1.99 | 1.06 | 0.37 | 3.84 | -0.19 |
| | | 4 | 5 | 0.59 | 0.29 | 0.36 | 1.09 | 2.35 | 1.05 | 1.27 | 3.67 | 0.03 |
| 11 | Overall | | 263 | 0.84 | 0.39 | 0.21 | 2.20 | 2.18 | 1.29 | -1.11 | 5.48 | -0.04 |
| | Item Type | SR | 32 | 0.45 | 0.21 | 0.21 | 1.22 | 1.05 | 1.27 | -1.06 | 3.63 | -0.43 |
| | | CR | 231 | 0.90 | 0.38 | 0.22 | 2.20 | 2.33 | 1.21 | -1.11 | 5.48 | -0.17 |
| | Score Categories | 2 | 213 | 0.90 | 0.40 | 0.21 | 2.20 | 2.28 | 1.29 | -1.06 | 5.48 | -0.10 |
| | | 3 | 41 | 0.62 | 0.23 | 0.22 | 1.19 | 1.78 | 1.19 | -0.87 | 4.45 | -0.16 |
| | | 4 | 9 | 0.59 | 0.24 | 0.35 | 1.01 | 1.45 | 1.15 | -1.11 | 2.51 | 0.20 |
| | Claim Area | 1 | 204 | 0.84 | 0.38 | 0.21 | 2.18 | 2.09 | 1.31 | -1.11 | 5.48 | -0.06 |
| | | 2 | 31 | 0.85 | 0.49 | 0.24 | 2.20 | 2.81 | 1.20 | -0.53 | 4.90 | -0.04 |
| | | 3 | 20 | 0.83 | 0.33 | 0.33 | 1.68 | 1.74 | 0.98 | 0.26 | 4.10 | -0.03 |
| | | 4 | 8 | 0.92 | 0.54 | 0.34 | 2.07 | 2.96 | 0.79 | 1.61 | 4.18 | 0.25 |

## Ability Estimates Evaluation

It is worthwhile to see how the ability estimates and scales vary among all three model combinations. It is reasonable to expect that the correlations of ability estimates will be very high across models, because for a given examinee, the same item responses are used for all three ability estimates; all that differs is 1) how the item responses are weighted, and 2) how the ability scales differ in terms of "stretching out" or "pushing in" various parts of the ability scale.[4]

Tables 19 and 20 summarize means and standard deviations of theta estimates and their correlations across different model combinations for ELA and Math, respectively. Figures B.3 and B.4 present scatter plots of theta estimates under different model choices for ELA and Math, respectively. Results show that the ability estimates across all three models are highly correlated. The scatter plots show that 2PL/GPC produced ability estimates that were most similar to the 3PL/GPC in the middle of the ability scale. Despite the difference between item parameter estimates produced by the 1PL/PC and the 3PL/GPC, the ability scale produced by the 1PL/PC is very similar to that produced by 3PL/GPC, and the two ability scales exhibit a linear relationship.

---

[4]The three models do produce different scales when applied to multiple-choice data where it is possible for very low ability students to correctly guess the keyed answer (Yen, 1981).

Table 19. ELA Correlations of Ability Estimates across Different Model Combinations

| Grade | Model | Theta Summary | | Theta Correlations | | |
|---|---|---|---|---|---|---|
| | | Mean | SD | 1PL/PC | 2PL/GPC | 3PL/GPC |
| 3 | 1PL/PC | -0.02 | 1.10 | 1.00 | 0.99 | 0.98 |
| | 2PL/GPC | -0.01 | 1.10 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.01 | 1.10 | | | 1.00 |
| 4 | 1PL/PC | -0.01 | 1.13 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | 0.01 | 1.14 | | 1.00 | 0.99 |
| | 3PL/GPC | 0.01 | 1.14 | | | 1.00 |
| 5 | 1PL/PC | -0.01 | 1.14 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | 0.00 | 1.16 | | 1.00 | 0.98 |
| | 3PL/GPC | 0.00 | 1.18 | | | 1.00 |
| 6 | 1PL/PC | -0.01 | 1.16 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | 0.00 | 1.18 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.01 | 1.19 | | | 1.00 |
| 7 | 1PL/PC | -0.01 | 1.16 | 1.00 | 0.97 | 0.95 |
| | 2PL/GPC | 0.01 | 1.19 | | 1.00 | 0.98 |
| | 3PL/GPC | -0.01 | 1.19 | | | 1.00 |
| 8 | 1PL/PC | -0.01 | 1.17 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | 0.00 | 1.19 | | 1.00 | 0.99 |
| | 3PL/GPC | 0.00 | 1.20 | | | 1.00 |
| 9 | 1PL/PC | -0.01 | 1.17 | 1.00 | 0.97 | 0.96 |
| | 2PL/GPC | 0.00 | 1.20 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.01 | 1.21 | | | 1.00 |
| 10 | 1PL/PC | -0.02 | 1.15 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | 0.00 | 1.15 | | 1.00 | 0.99 |
| | 3PL/GPC | 0.00 | 1.15 | | | 1.00 |
| 11 | 1PL/PC | -0.02 | 1.12 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | -0.03 | 1.14 | | 1.00 | 0.98 |
| | 3PL/GPC | -0.04 | 1.15 | | | 1.00 |

Table 20. Math Correlations of Ability Estimates across Different Model Combinations

| Grade | Model | Theta Summary | | Theta Correlations | | |
|---|---|---|---|---|---|---|
| | | Mean | SD | 1PL/PC | 2PL/GPC | 3PL/GPC |
| 3 | 1PL/PC | -0.01 | 1.10 | 1.00 | 0.99 | 0.98 |
| | 2PL/GPC | -0.03 | 1.11 | | 1.00 | 1.00 |
| | 3PL/GPC | -0.03 | 1.11 | | | 1.00 |
| 4 | 1PL/PC | -0.01 | 1.07 | 1.00 | 0.99 | 0.98 |
| | 2PL/GPC | -0.04 | 1.09 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.06 | 1.06 | | | 1.00 |
| 5 | 1PL/PC | -0.02 | 1.09 | 1.00 | 0.99 | 0.97 |
| | 2PL/GPC | -0.04 | 1.11 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.05 | 1.11 | | | 1.00 |
| 6 | 1PL/PC | 0.01 | 1.09 | 1.00 | 0.98 | 0.97 |
| | 2PL/GPC | -0.01 | 1.11 | | 1.00 | 0.99 |
| | 3PL/GPC | 0.00 | 1.09 | | | 1.00 |
| 7 | 1PL/PC | 0.00 | 1.09 | 1.00 | 0.98 | 0.96 |
| | 2PL/GPC | -0.02 | 1.11 | | 1.00 | 0.98 |
| | 3PL/GPC | -0.05 | 1.06 | | | 1.00 |
| 8 | 1PL/PC | 0.01 | 1.09 | 1.00 | 0.97 | 0.96 |
| | 2PL/GPC | -0.01 | 1.12 | | 1.00 | 0.99 |
| | 3PL/GPC | -0.01 | 1.11 | | | 1.00 |
| 9 | 1PL/PC | 0.00 | 1.14 | 1.00 | 0.95 | 0.92 |
| | 2PL/GPC | -0.07 | 1.16 | | 1.00 | 0.96 |
| | 3PL/GPC | -0.15 | 1.13 | | | 1.00 |
| 10 | 1PL/PC | -0.02 | 1.14 | 1.00 | 0.97 | 0.93 |
| | 2PL/GPC | -0.09 | 1.13 | | 1.00 | 0.97 |
| | 3PL/GPC | -0.27 | 1.03 | | | 1.00 |
| 11 | 1PL/PC | 0.06 | 1.01 | 1.00 | 0.95 | 0.92 |
| | 2PL/GPC | -0.08 | 1.01 | | 1.00 | 0.98 |
| | 3PL/GPC | -0.08 | 0.95 | | | 1.00 |

## 2.4 IRT Model Recommendation

Based on the model comparison analysis results, ETS recommends that the 2PL/GPC model be adopted as the IRT model combination for calibrating Smarter Balanced items and establishing a vertical scale. The 2PL/GPC model provides flexibility for estimating a range of item discriminations without the complications of implementing a 3PL/GPC model. The major limitation of the 2PL/GPC model in this setting is that it has not been previously used for vertical scaling in K–12 assessments.

This recommendation should be evaluated with caution given the experimental nature of the Pilot data, the possible change of item format from Pilot to Field Test to operational administration, and the lack of information about vertical scaling results for the three models.

## References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In (Eds.) Petrov, B.N. & Csaki. F. *Proceedings 2nd International Symposium Information Theory*, (pp.267–281), Budapest, Hungary: Akademia Kiado.

Bock, R. D. & Zimowski, M. F. (1997). Multiple Group IRT. In W.J. van der Linden and R. K. Hambleton (Ed.), *Handbook of Modern Item Response Theory* (pp. 433–448). New York: Springer-Verlag.

Briggs, D. C., & Weeks, J. P. (2009). The sensitivity of value-added modeling to the creation of a vertical score scale. *Education Finance and Policy, 4(4),* 384–414.

Dorans, N. J., & Kulick, E. (1983). *Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach* (RR-83-9). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23,* 355–368.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (RR-91-47). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Schmitt, A. P. (1993). Constructed response and differential item functioning: A pragmatic perspective. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 135–165). Hillsdale NJ: Erlbaum.

Fitzpatrick, A. R.; Link, V.; Yen, W. M.; Burket, G. R.; Ito, K.; & Sykes, R. (1996). Scaling performance assessments: A comparison of one-parameter and two-parameter Partial Credit Models. *Journal of Educational Measurement, 33,* 291–314.

Holland, P. W., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mantel, N., & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719–748.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

McKinley, R. L., & Reckase, M. D. (1983). An application of a multidimensional extension of the two-parameter logistic latent trait model (ONR-83-3). (ERIC Document Reproduction Service No. ED 240 168)

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Reckase, M. D.; Martineau, J. A.; & Kim, J.P. (2000, July). A vector approach to determining the number of dimensions needed to represent a set of variables. Paper presented at the annual meeting of the Psychometric Society, Vancouver, Canada.

Reckase, M. D.; Ackerman, T.A.; & Carlson, J.E. (1988). Building unidimensional tests using multidimensional items. *Journal of Educational Measurement, 25,* 193–203.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Sykes, R. C., and Yen, W. M. (2000). The scaling of mixed-item format tests with the one-parameter and two-parameter partial credit models. *Journal of Educational Measurement, 37(3),* 221–244.

Weeks. J. P. (2010). **Plink**: An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. Journal of Statistical Software, 35(12), 1–33. URL http://www.jstatsoft.org/v35/i12/.

Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory*. [Computer software]. Monterey, CA: Defense Manpower Data Center.

Yao, L., & Schwarz, R.D. (2006). A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests. *Applied Psychological Measurement, 30,* 469–492.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5,* 245–262.

Yen, W. M. (1993). Scaling performance assessments: strategies for managing local item dependence*. Journal of Educational Measurement*, 30, 187–214.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R.L. Brennan (eds.), *Educational Measurement (Fourth Edition),* Westport, CT: American Council on Education and Praeger Publishing.

Zwick, R.; Donoghue, J. R.; & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30(3),* 233–251.

# Appendix A: Item Vector Plots
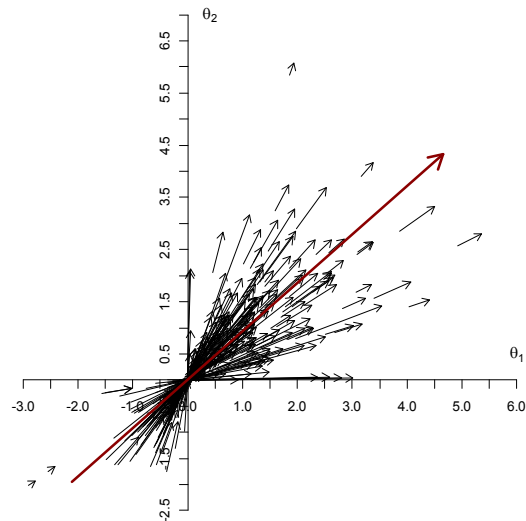


Figure A.1. Item Vector Plot for ELA Grade 3 (Within Grade)
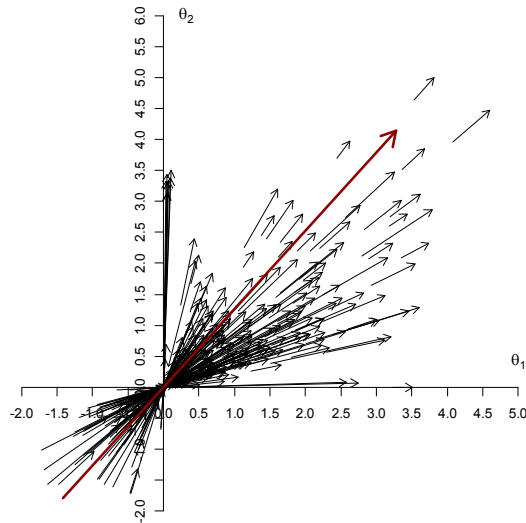


Figure A.2. Item Vector Plot for ELA Grade 4 (Within Grade)
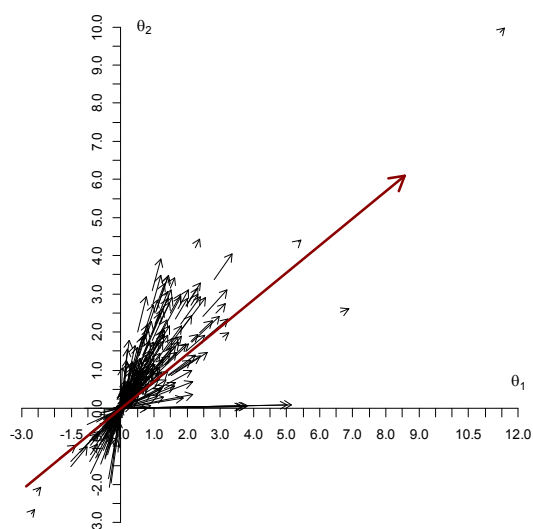
Figure A.3. Item Vector Plot for ELA Grade 5 (Within Grade)
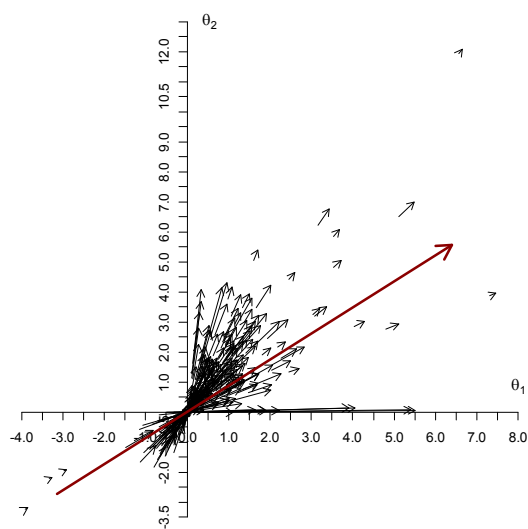


Figure A.4. Item Vector Plot for ELA Grade 6 (Within Grade)
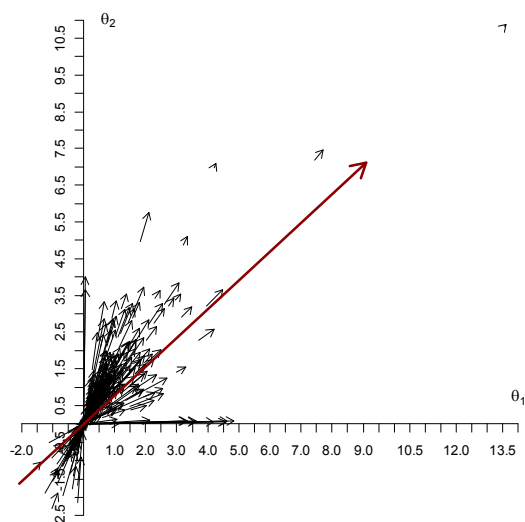
Figure A.5. Item Vector Plot for ELA Grade 7 (Within Grade)
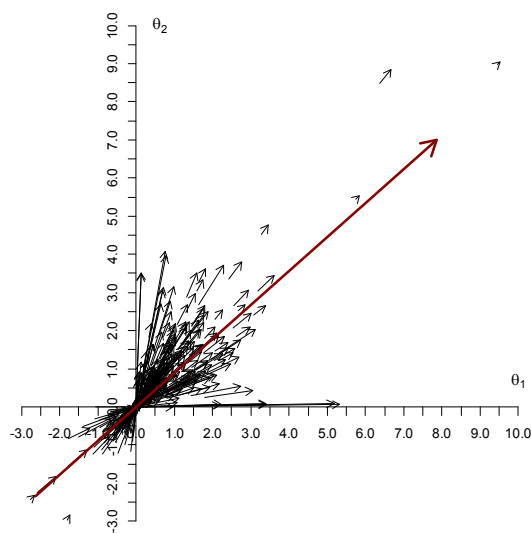


Figure A.6. Item Vector Plot for ELA Grade 8 (Within Grade)
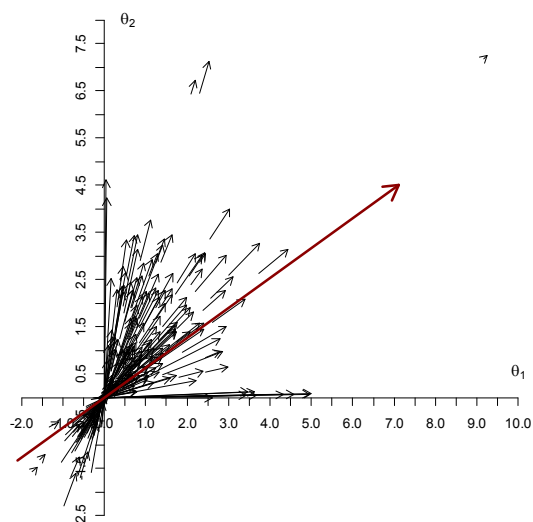
Figure A.7. Item Vector Plot for ELA Grade 9 (Within Grade)
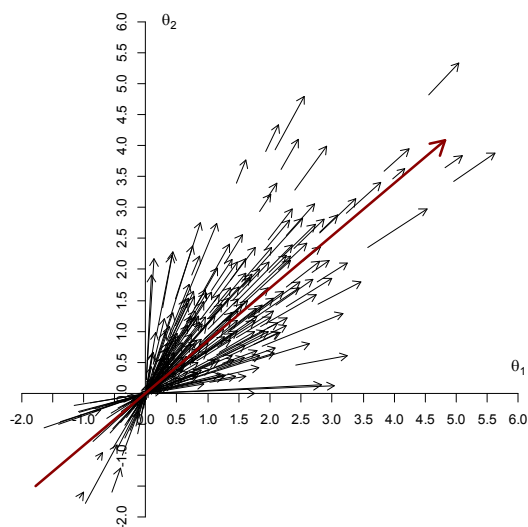


Figure A.8. Item Vector Plot for ELA Grade 10 (Within Grade)
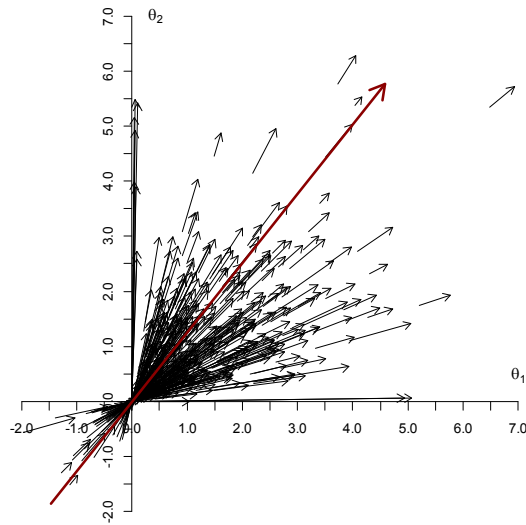
Figure A.9. Item Vector Plot for ELA Grade 11 (Within Grade)
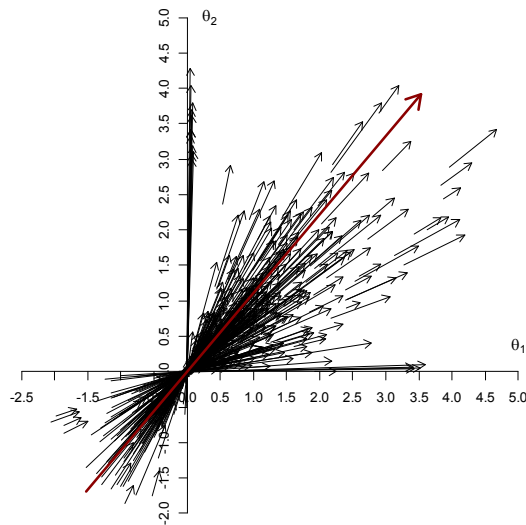


Figure A.10. Item Vector Plot for ELA Grades 3 and 4 (Across Grades)
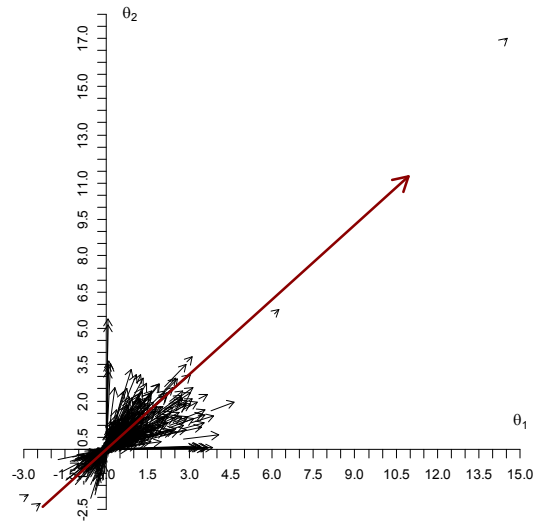
Figure A.11. Item Vector Plot for ELA Grades 4 and 5 (Across Grades)
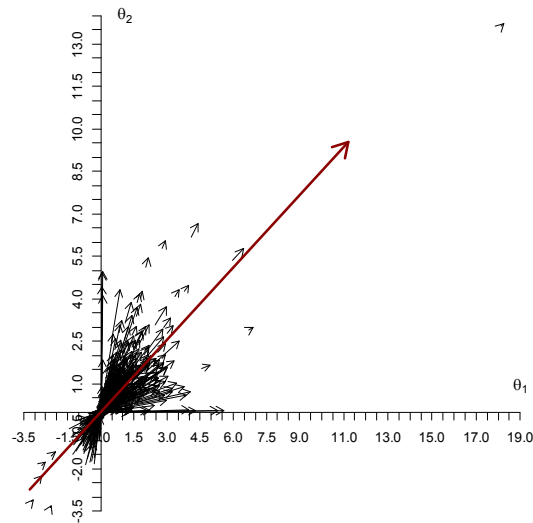


Figure A.12. Item Vector Plot for ELA Grades 5 and 6 (Across Grades)
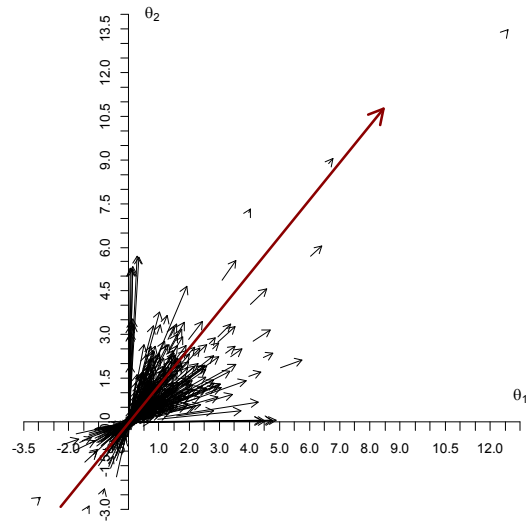
Figure A.13. Item Vector Plot for ELA Grades 6 and 7 (Across Grades)
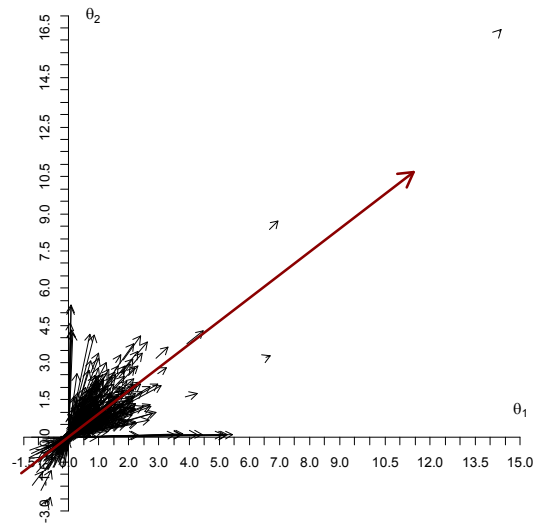


Figure A.14. Item Vector Plot for ELA Grades 7 and 8 (Across Grades)
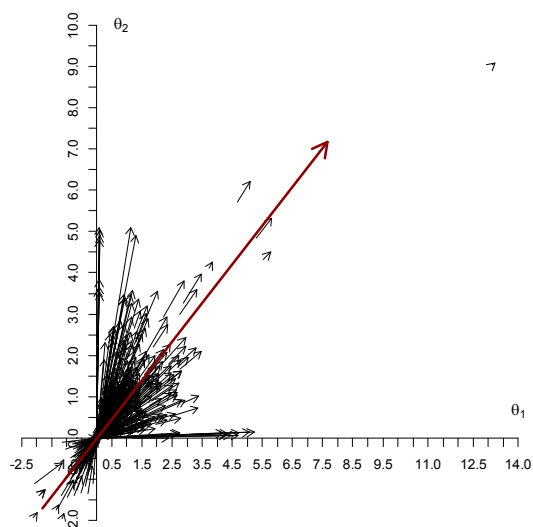
Figure A.15. Item Vector Plot for ELA Grades 8 and 9 (Across Grades)
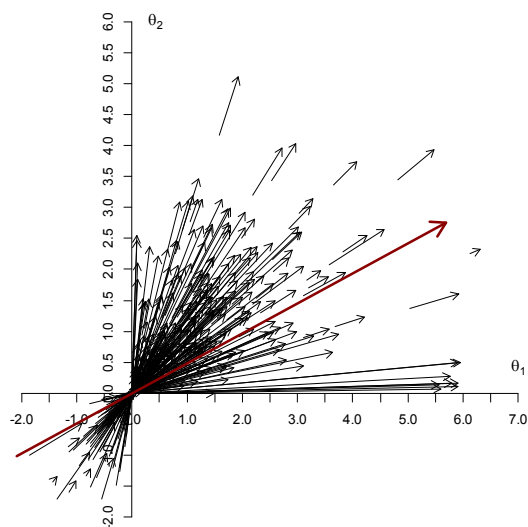


Figure A.16. Item Vector Plot for ELA Grades 9 and 10 (Across Grades)
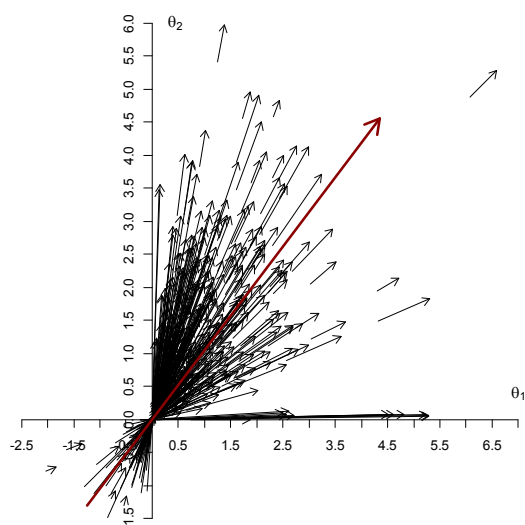
Figure A.17. Item Vector Plot for ELA Grades 10 and 11 (Across Grades)
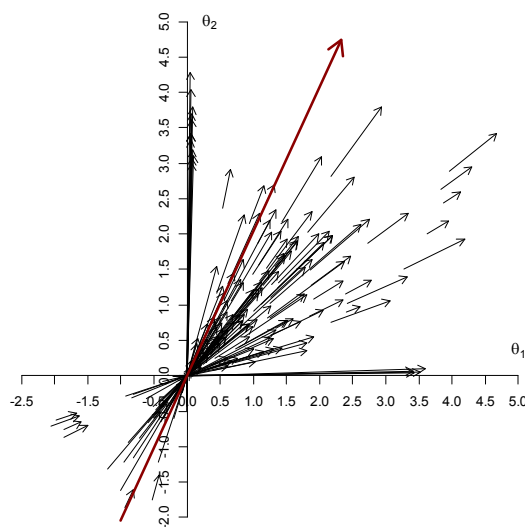


Figure A.18. Item Vector Plots for the Subset of ELA Grade 3 and 4 Vertical Linking Items
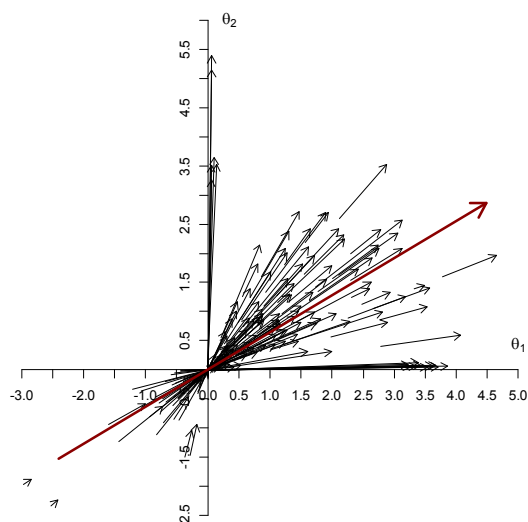
Figure A.19. Item Vector Plots for the Subset of ELA Grade 4 and 5 Vertical Linking Items
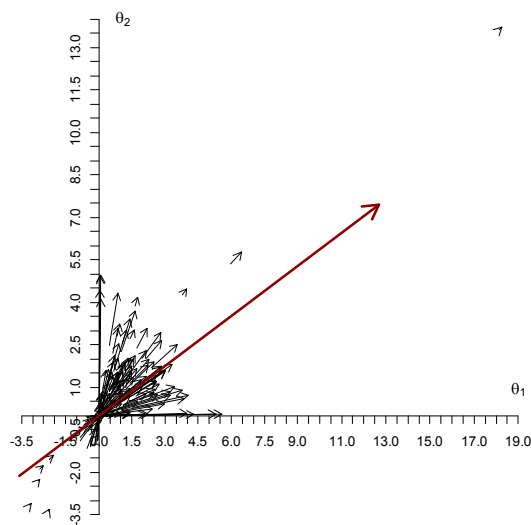


Figure A.20. Item Vector Plots for the Subset of ELA Grade 5 and 6 Vertical Linking Items
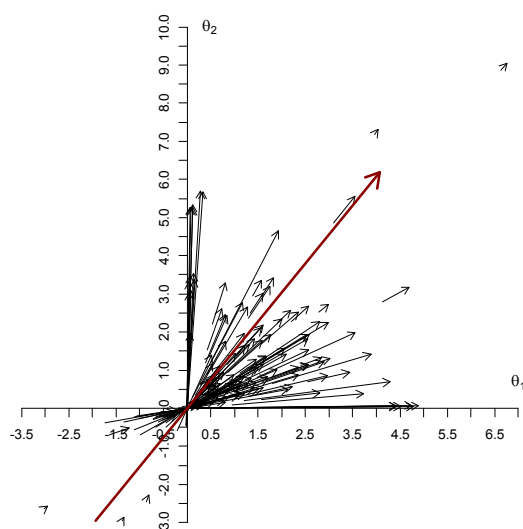
Figure A.21. Item Vector Plots for the Subset of ELA Grade 6 and 7 Vertical Linking Items
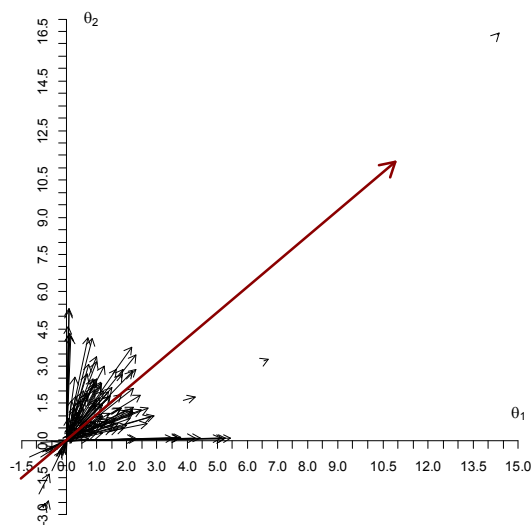


Figure A.22. Item Vector Plots for the Subset of ELA Grade 7 and 8 Vertical Linking Items
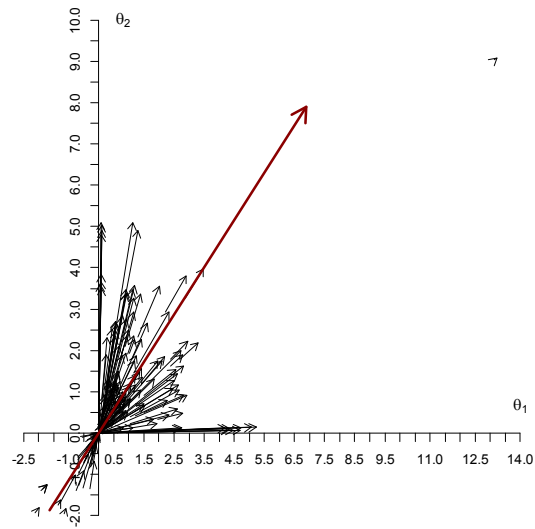
Figure A.23. Item Vector Plots for the Subset of ELA Grade 8 and 9 Vertical Linking Items
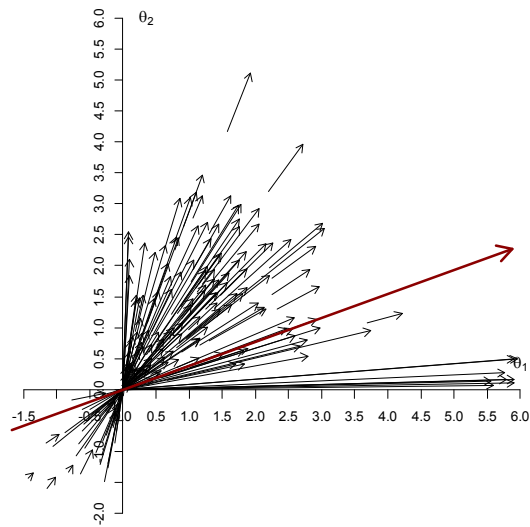


Figure A.24. Item Vector Plots for the Subset of ELA Grade 9 and 10 Vertical Linking Items
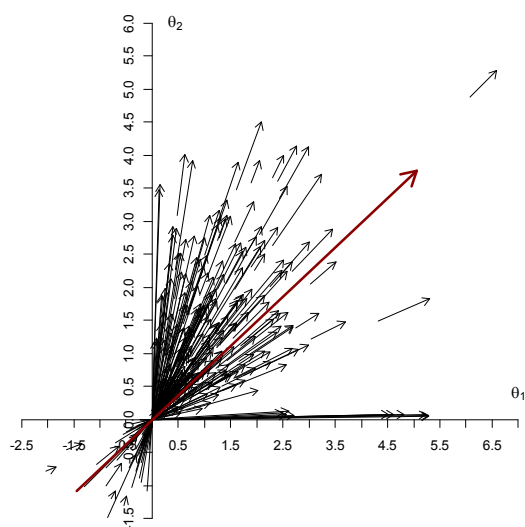
Figure A.25. Item Vector Plots for the Subset of ELA Grade 10 and 11 Vertical Linking Items
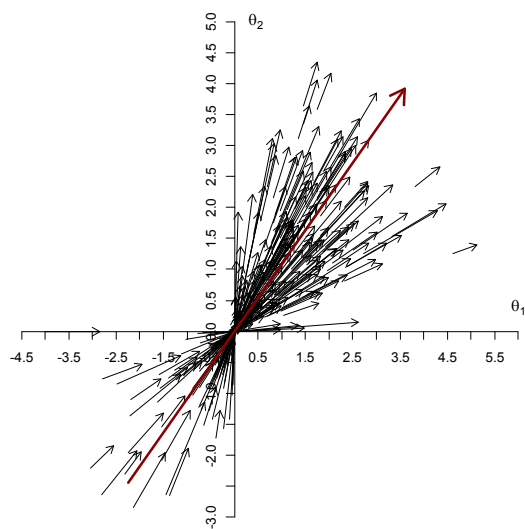


Figure A.26. Item Vector Plot for Math Grade 3 (Within Grade)
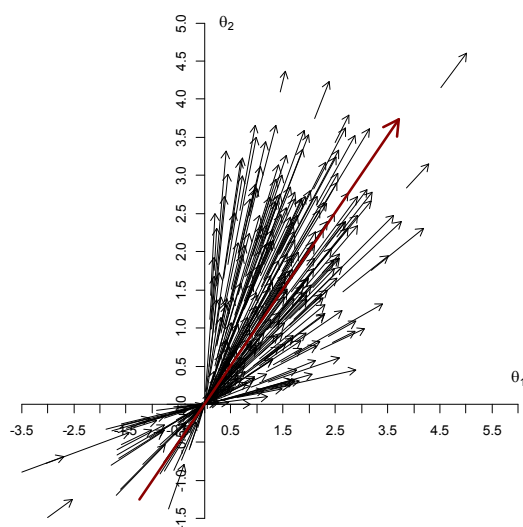
Figure A.27. Item Vector Plot for Math Grade 4 (Within Grade)
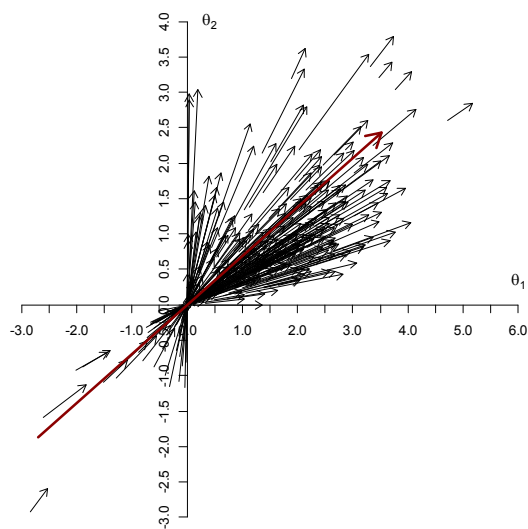


Figure A.28. Item Vector Plot for Math Grade 5 (Within Grade)
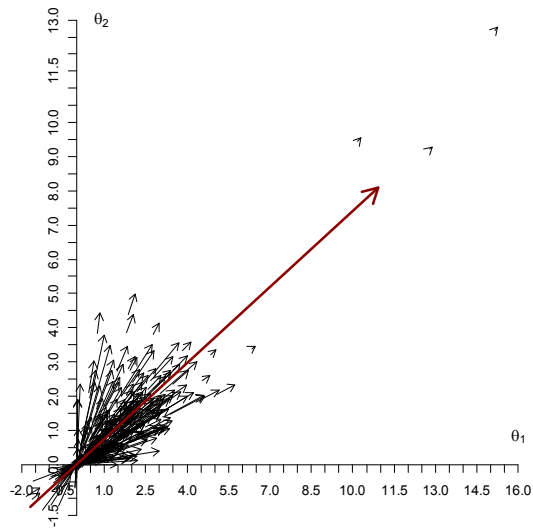
Figure A.29. Item Vector Plot for Math Grade 6 (Within Grade)
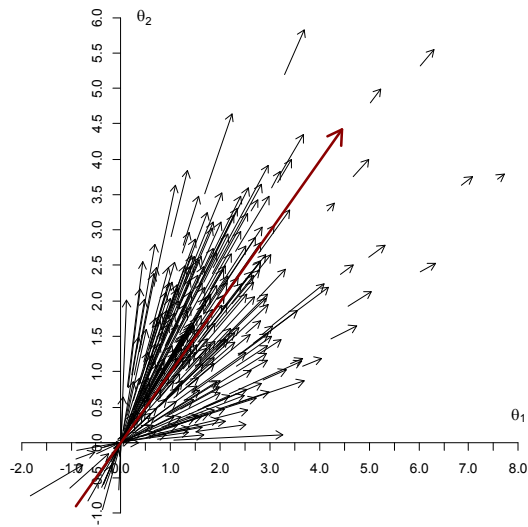


Figure A.30. Item Vector Plot for Math Grade 7 (Within Grade)
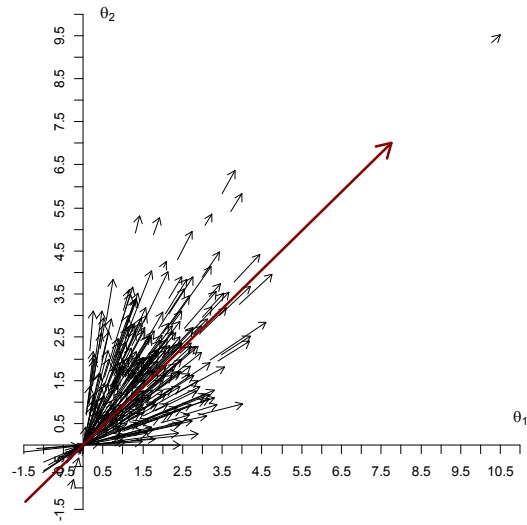
Figure A.31. Item Vector Plot for Math Grade 8 (Within Grade)
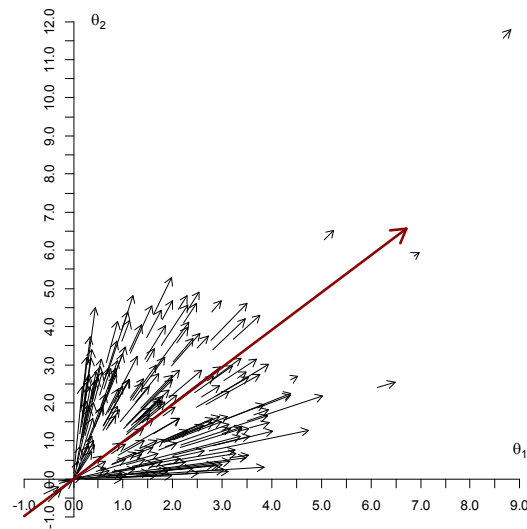


Figure A.32. Item Vector Plot for Math Grade 9 (Within Grade)
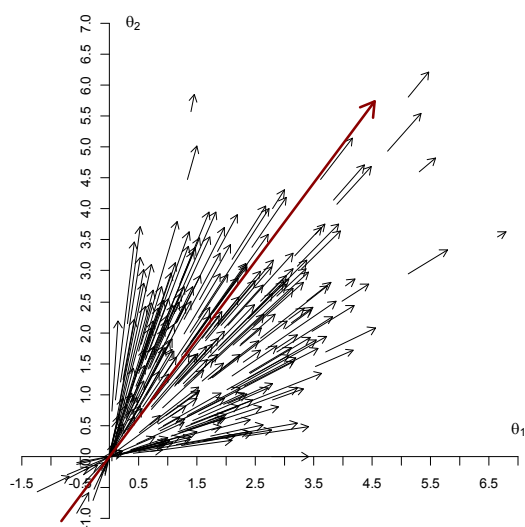
Figure A.33. Item Vector Plot for Math Grade 10 (Within Grade)
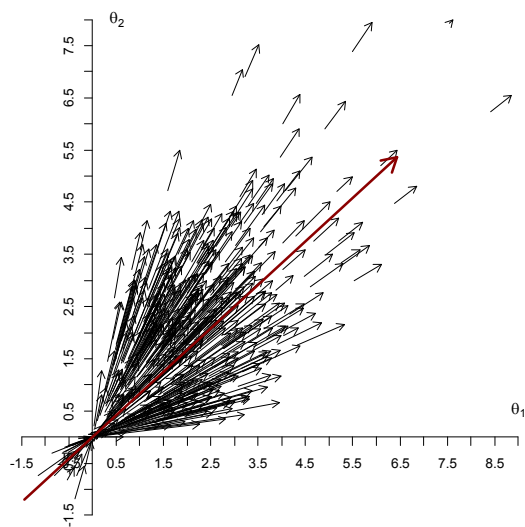


Figure A.34. Item Vector Plot for Math Grade 11 (Within Grade)
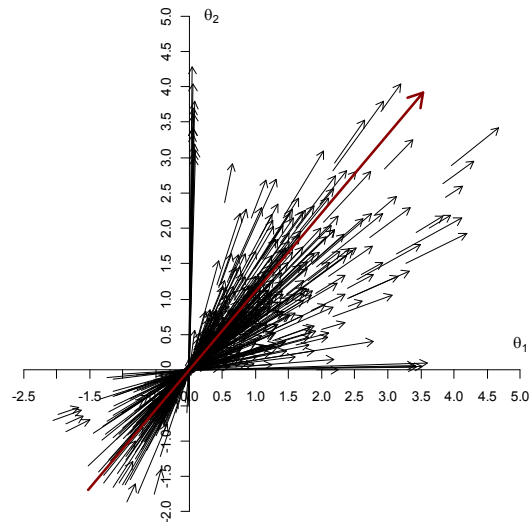
Figure A.35. Item Vector Plot for Math Grades 3 and 4 (Across Grades)
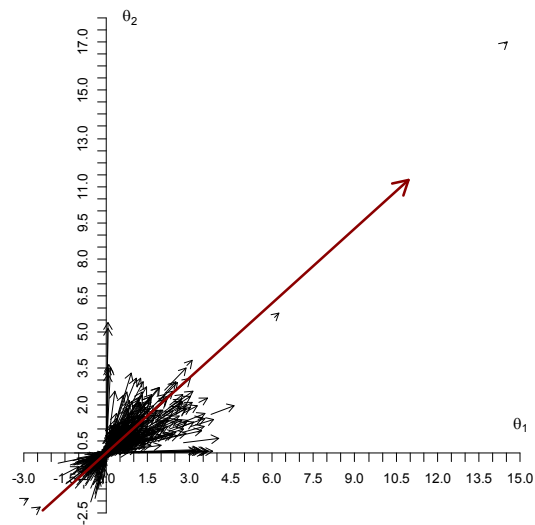


Figure A.36. Item Vector Plot for Math Grades 4 and 5 (Across Grades)
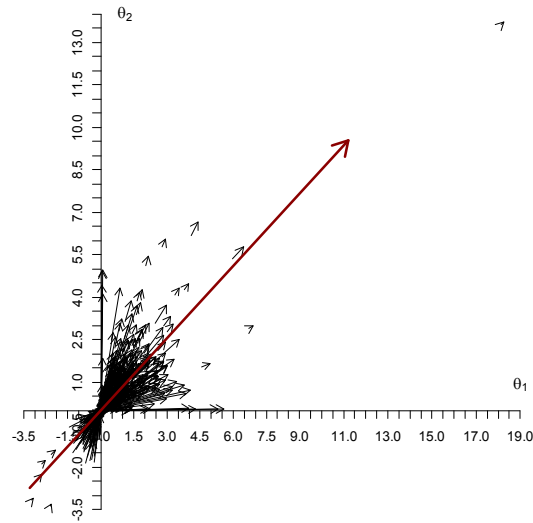
Figure A.37. Item Vector Plot for Math Grades 5 and 6 (Across Grades)
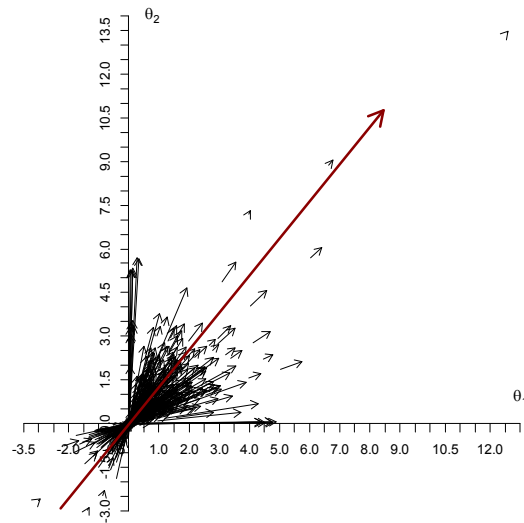


Figure A.38. Item Vector Plot for Math Grades 6 and 7 (Across Grades)
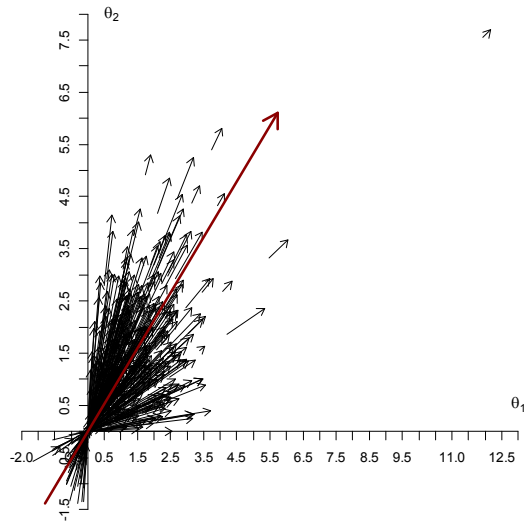
Figure A.39. Item Vector Plot for Math Grades 7 and 8 (Across Grades)
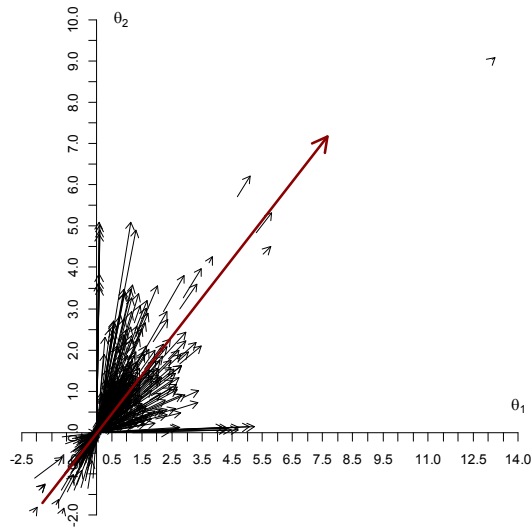


Figure A.40. Item Vector Plot for Math Grades 8 and 9 (Across Grades)
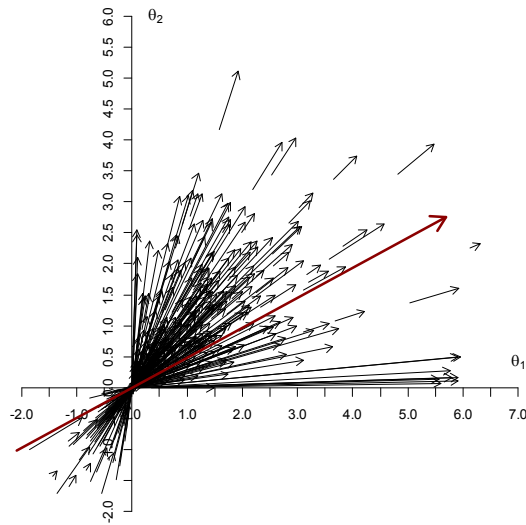
Figure A.41. Item Vector Plot for Math Grades 9 and 10 (Across Grades)
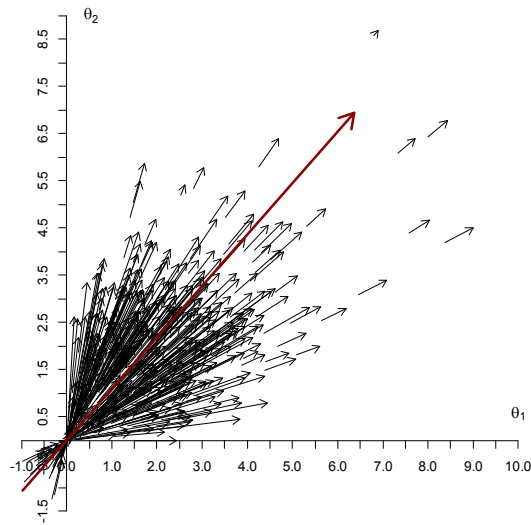


Figure A.42. Item Vector Plot for Math Grades 10 and 11 (Across Grades)
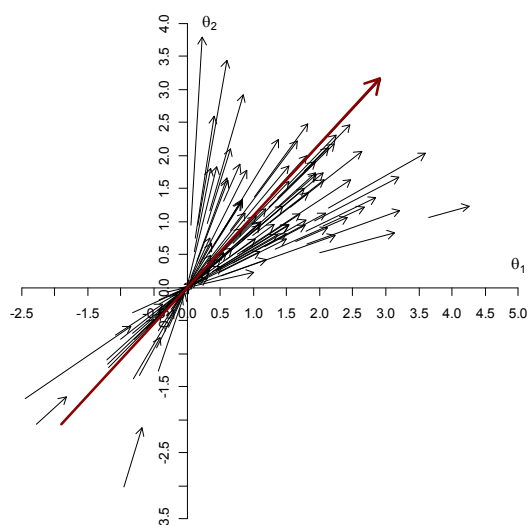
Figure A.43. Item Vector Plot for the Subset of Math Grade 3 and 4 Vertical Linking Items
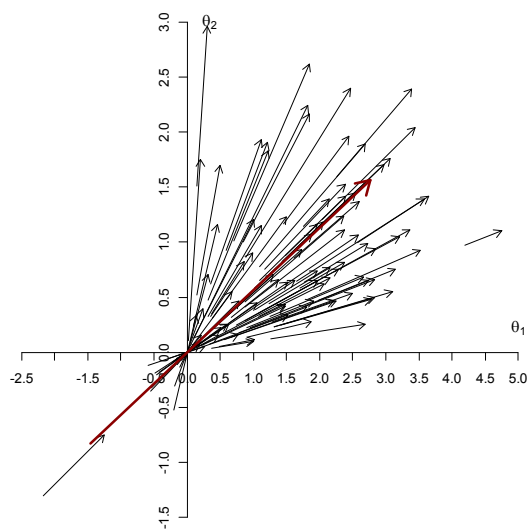


Figure A.44. Item Vector Plots for the Subset of Math Grade 4 and 5 Vertical Linking Items
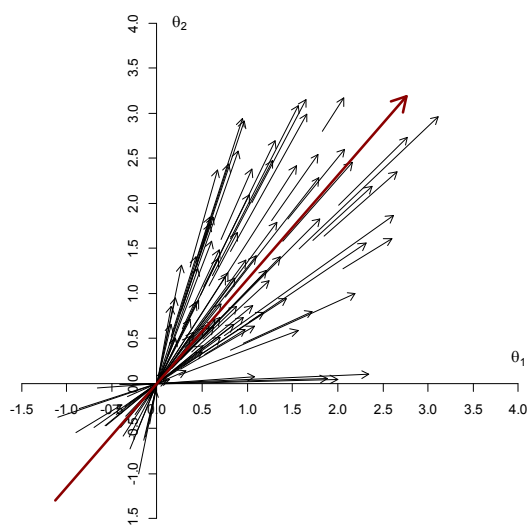
Figure A.45. Item Vector Plots for the Subset of Math Grade 5 and 6 Vertical Linking Items
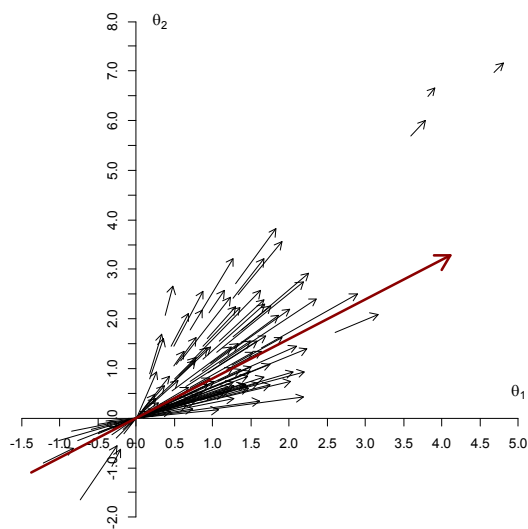


Figure A.46. Item Vector Plots for the Subset of Math Grade 6 and 7 Vertical Linking Items
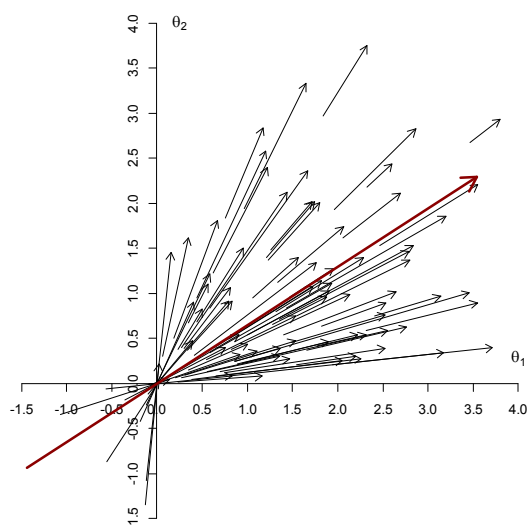
Figure A.47. Item Vector Plots for the Subset of Math Grade 7 and 8 Vertical Linking Items
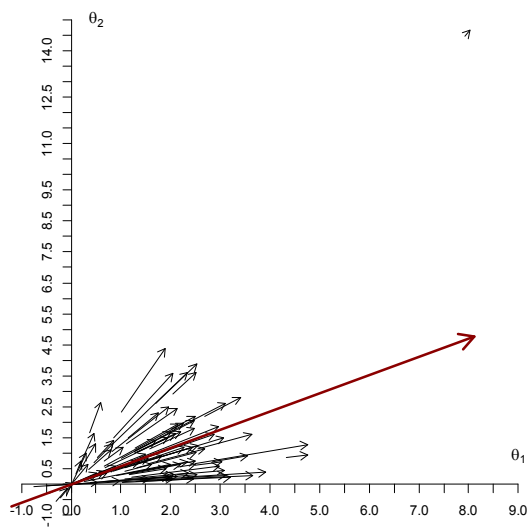


Figure A.48. Item Vector Plots for the Subset of Math Grade 8 and 9 Vertical Linking Items
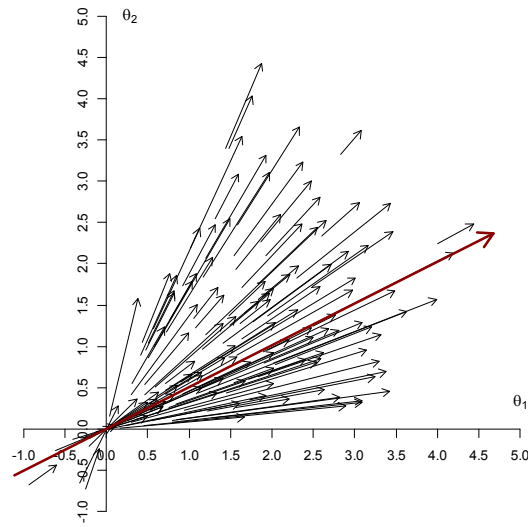
Figure A.49. Item Vector Plots for the Subset of Math Grade 9 and 10 Vertical Linking Items
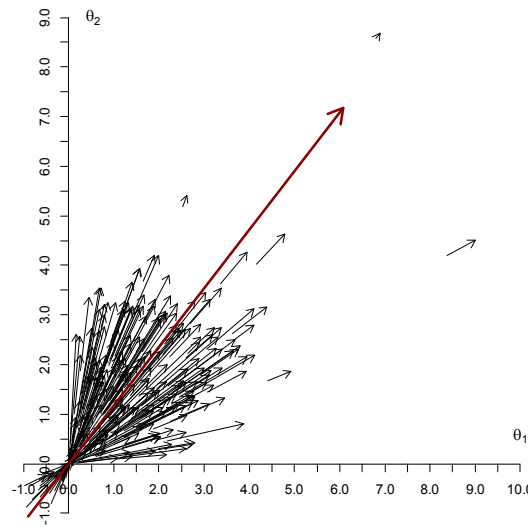


Figure A.50. Item Vector Plots for the Subset of Math Grade 10 and 11 Vertical Linking Items

### Appendix B: Tables and Figures IRT Model Comparison

Figure B.1. Scatter Plot of ELA 2PL/GPC Slope and Difficulty Estimates by Item Type, Score Category and Claim



**ELA G03 Scatter Plot of 2PL/GPC *a* and *b* by Item Type**

**ELA G03 Scatter Plot of 2PL/GPC *a* and *b* by Score Category**

**ELA G03 Scatter Plot of 2PL/GPC *a* and *b* by Claim**

**ELA G04 Scatter Plot of 2PL/GPC *a* and *b* by Item Type**

**ELA G04 Scatter Plot of 2PL/GPC *a* and *b* by Score Category**

**ELA G04 Scatter Plot of 2PL/GPC *a* and *b* by Claim**
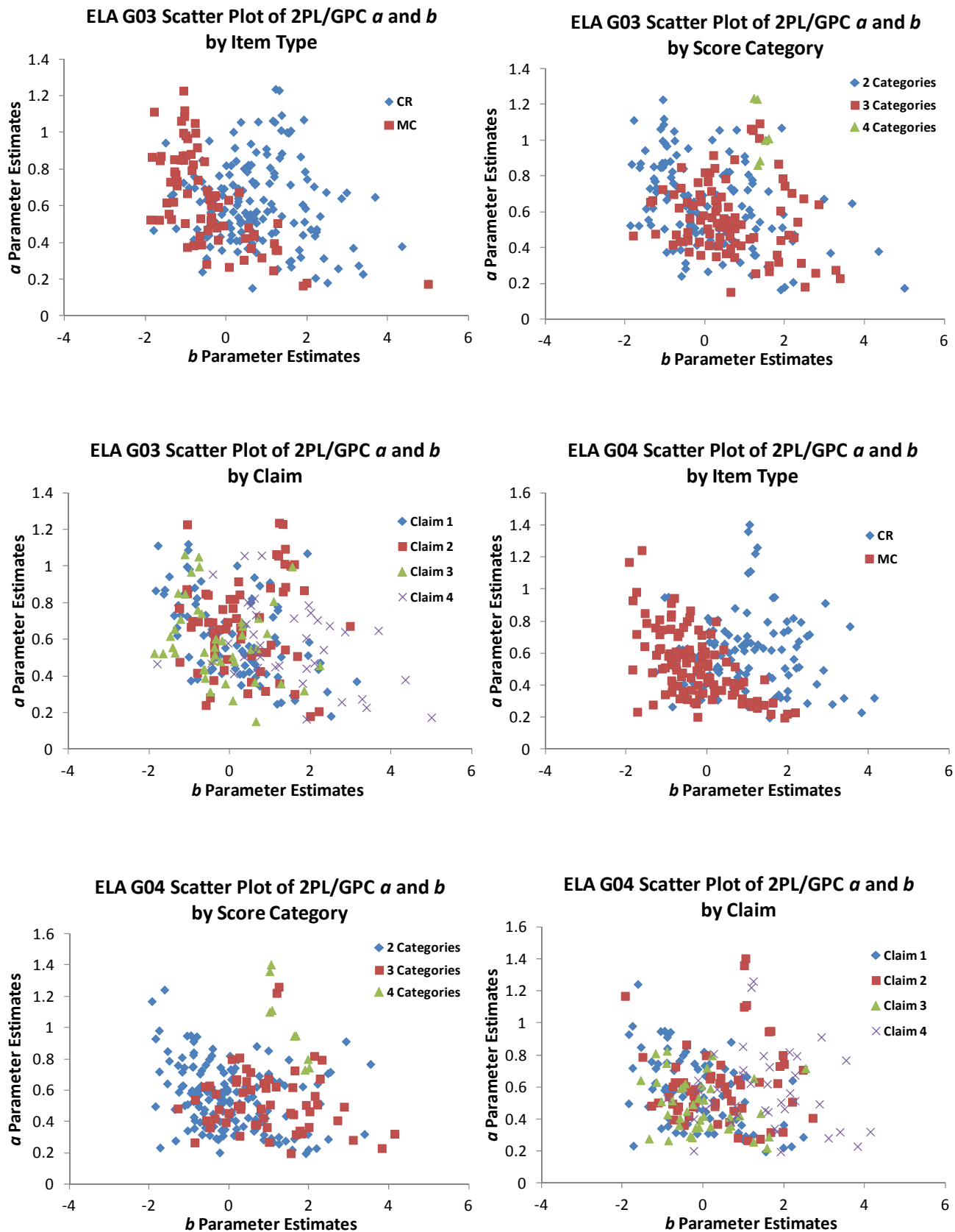
ELA G05 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
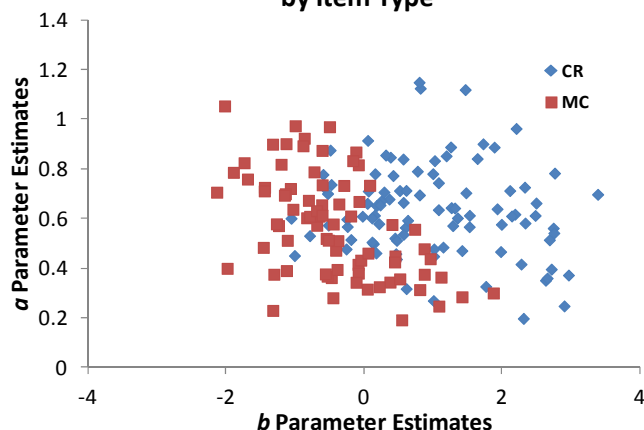
ELA G05 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
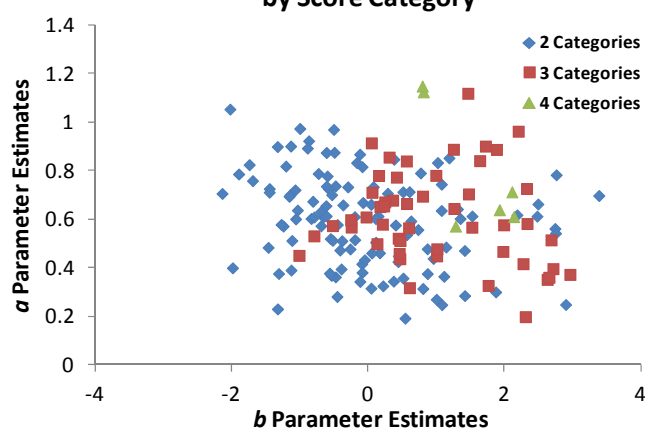
ELA G05 Scatter Plot of 2PL/GPC *a* and *b* by Claim
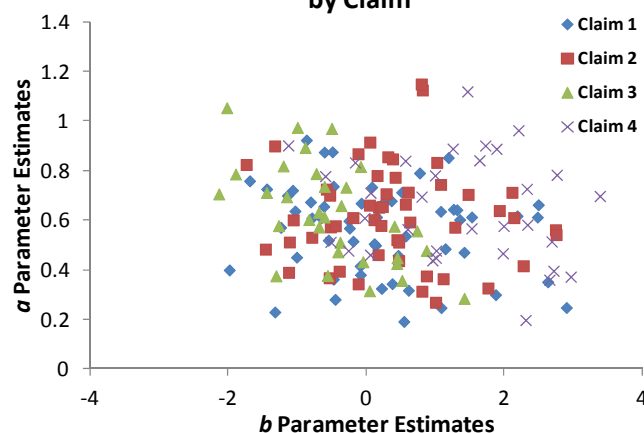
ELA G06 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
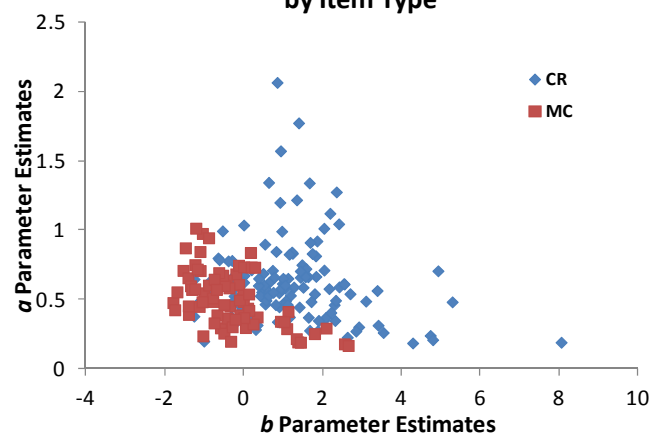
ELA G06 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
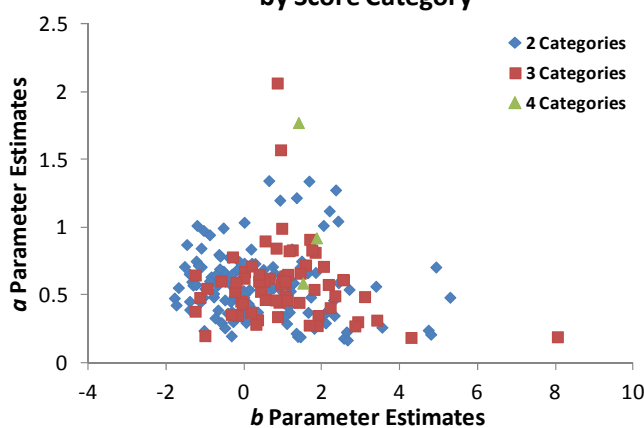
ELA G06 Scatter Plot of 2PL/GPC *a* and *b* by Claim
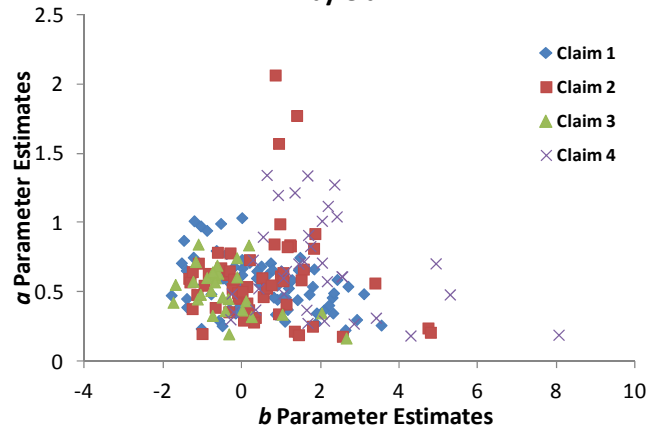
ELA G07 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
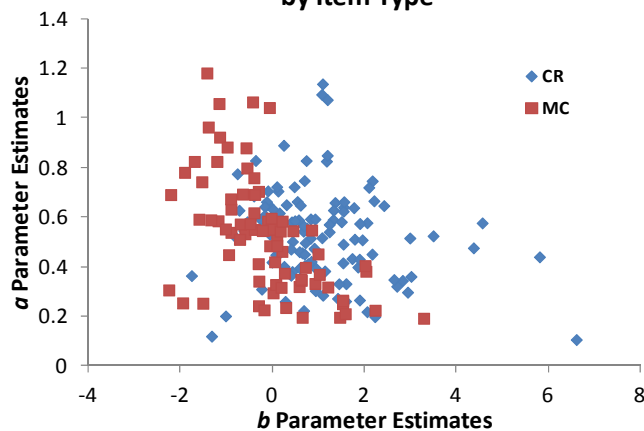
ELA G07 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
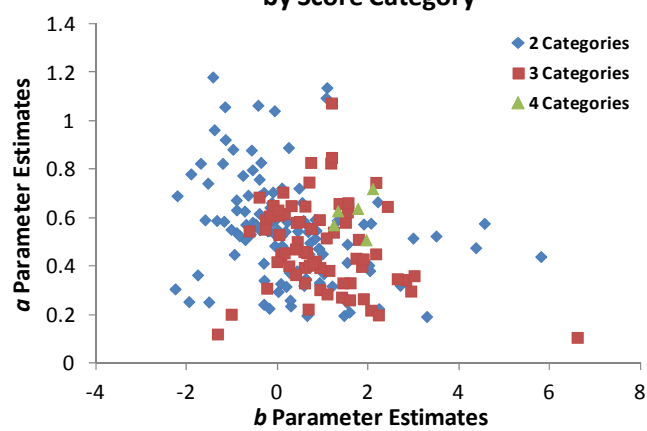
ELA G07 Scatter Plot of 2PL/GPC *a* and *b* by Claim
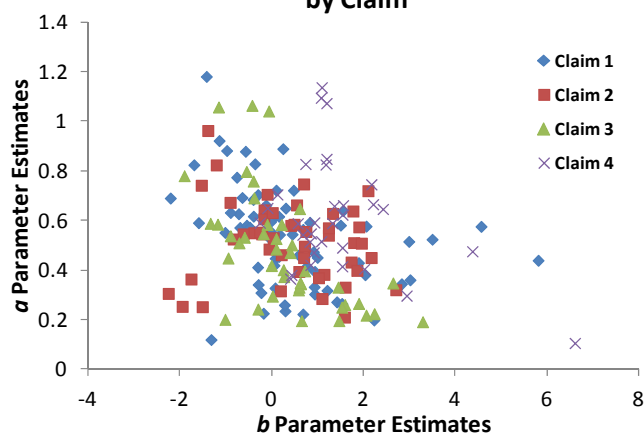
ELA G08 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
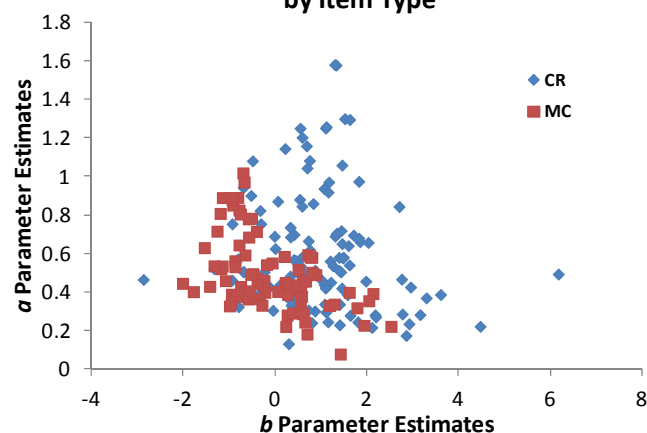
ELA G08 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
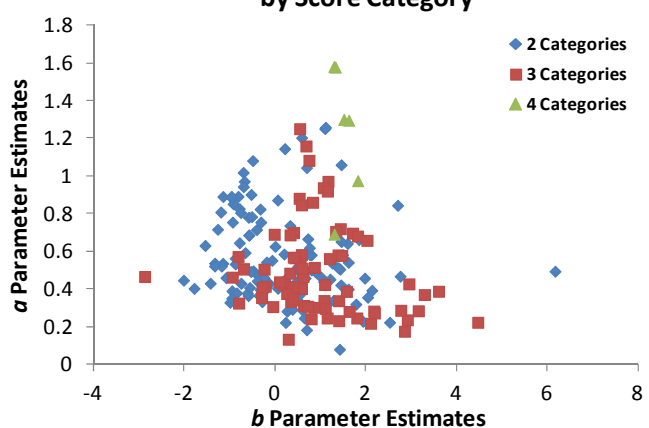
ELA G08 Scatter Plot of 2PL/GPC *a* and *b* by Claim
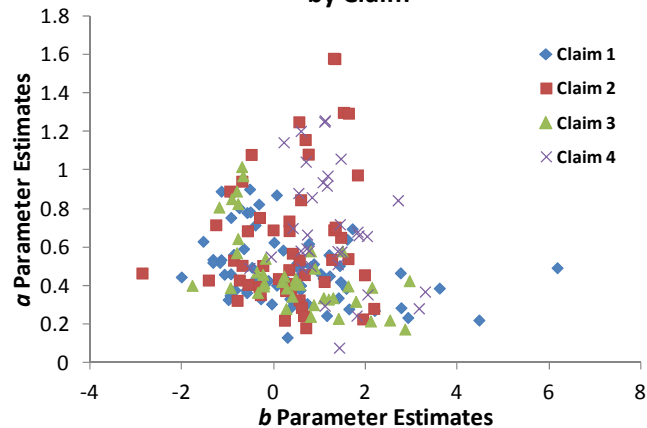
ELA G09 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
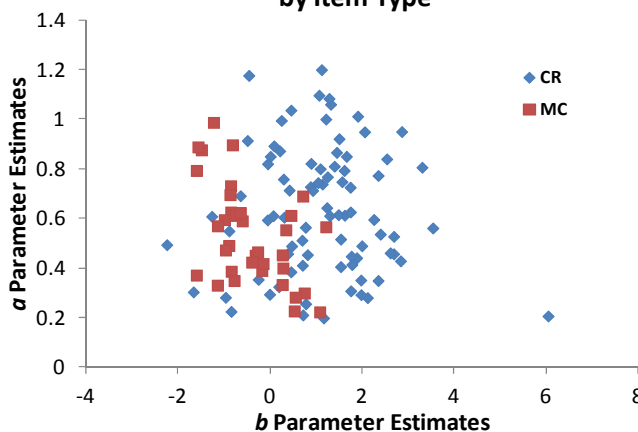
ELA G09 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
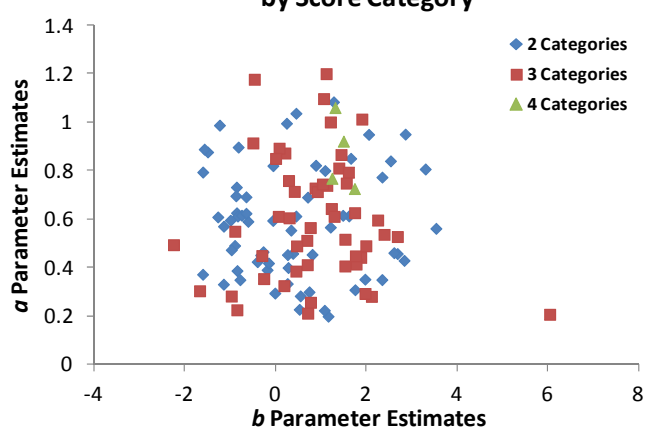
ELA G09 Scatter Plot of 2PL/GPC *a* and *b* by Claim
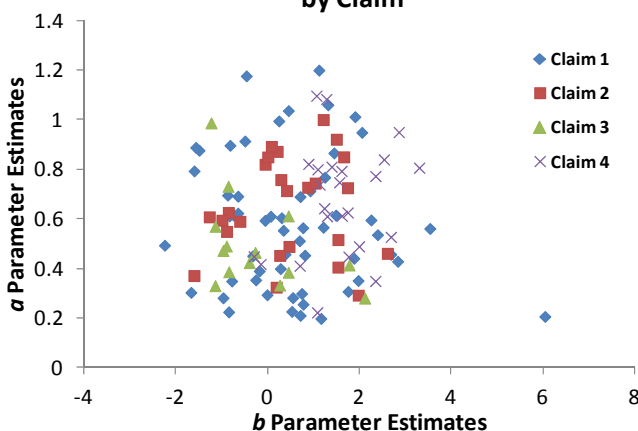
ELA G10 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
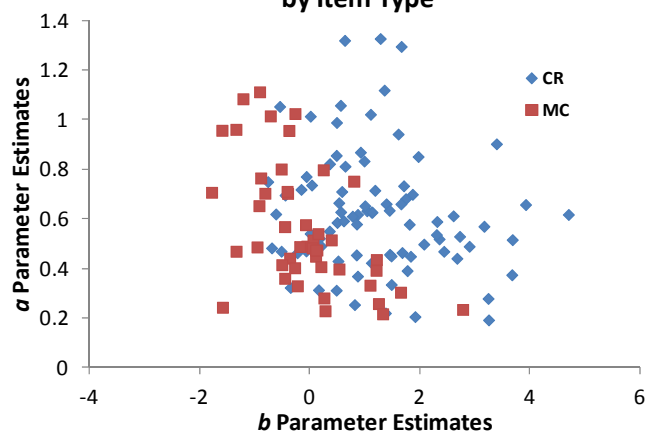
ELA G10 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
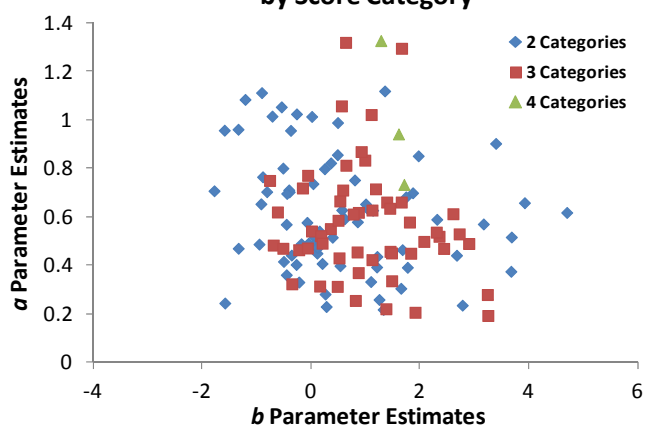
ELA G10 Scatter Plot of 2PL/GPC *a* and *b* by Claim
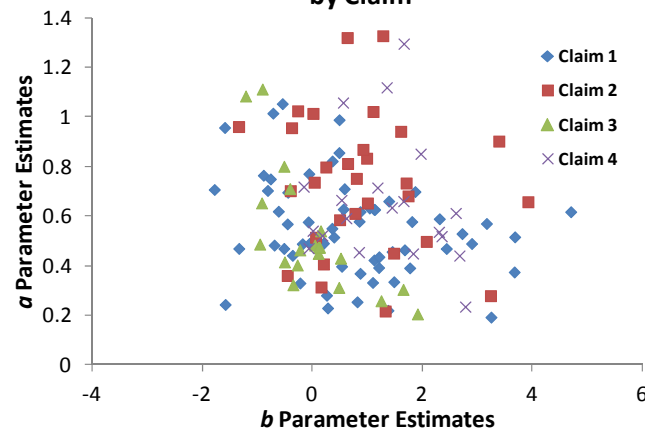
ELA G11 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
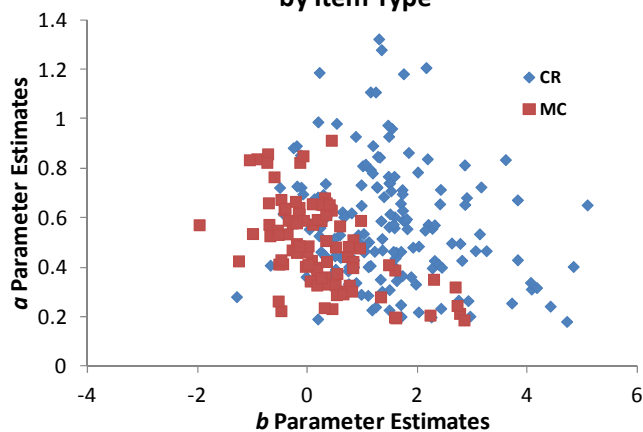
ELA G11 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
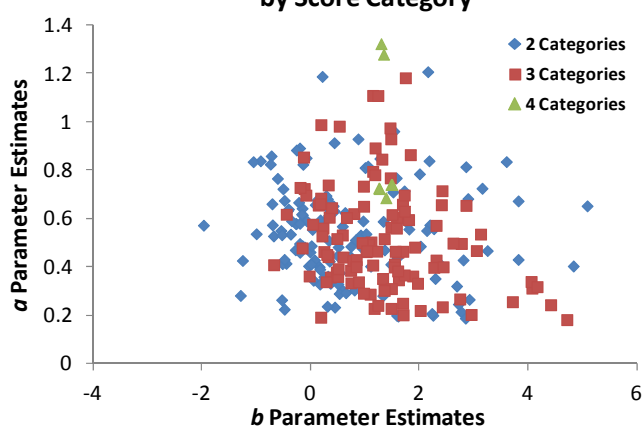
ELA G11 Scatter Plot of 2PL/GPC *a* and *b* by Claim
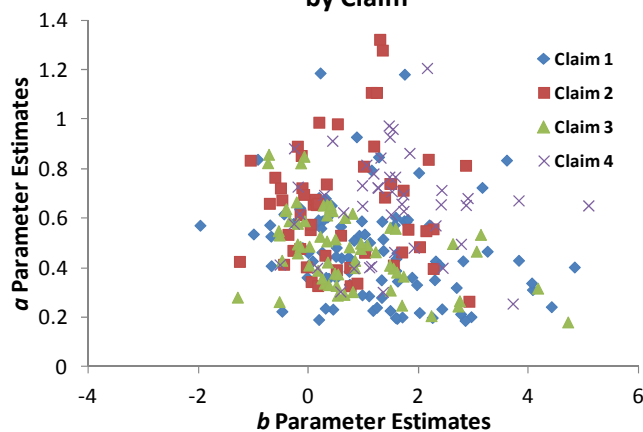
Figure B.2. Scatter Plot of Math 2PL/GPC Slope and Difficulty Estimates by Item Type, Score Category and Claim
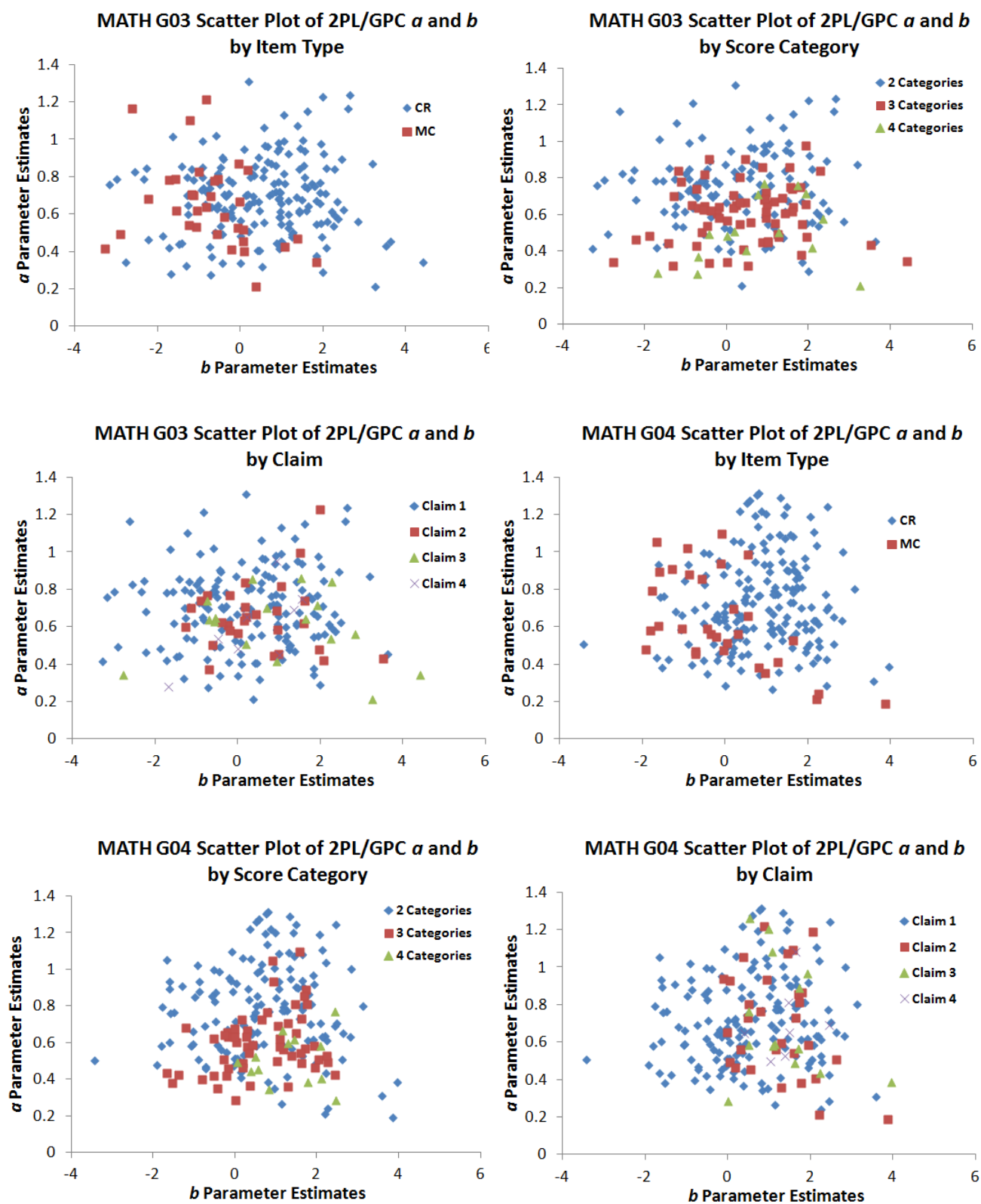
MATH G05 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
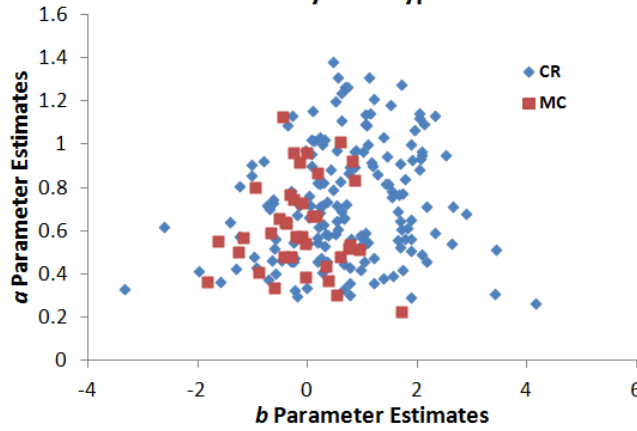
MATH G05 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
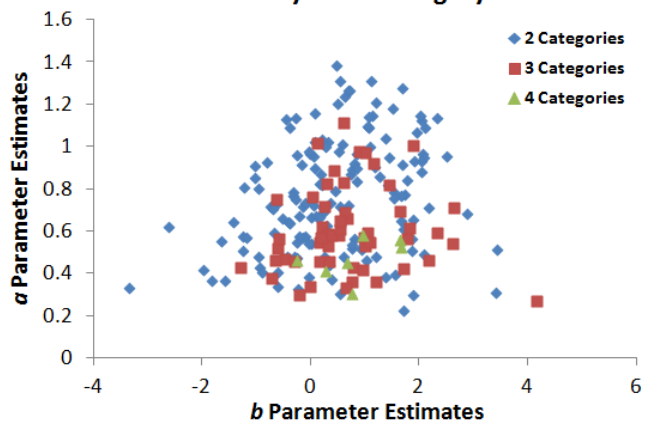
MATH G05 Scatter Plot of 2PL/GPC *a* and *b* by Claim
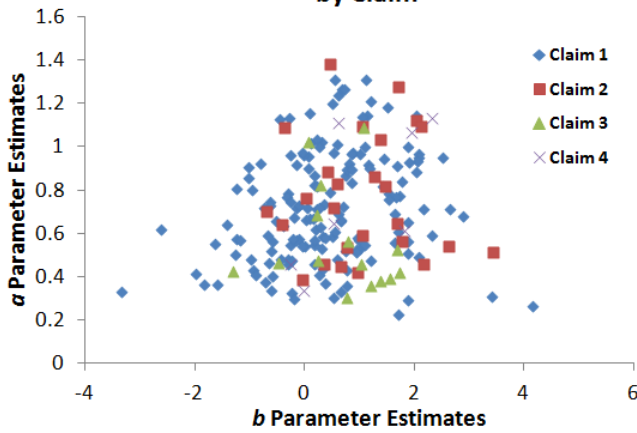
MATH G06 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
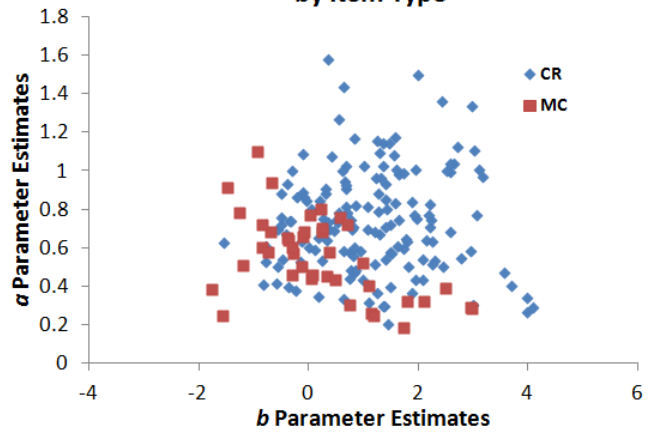
MATH G06 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
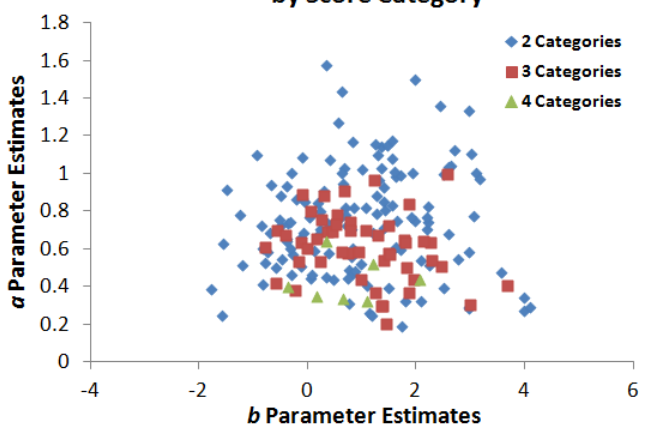
MATH G06 Scatter Plot of 2PL/GPC *a* and *b* by Claim
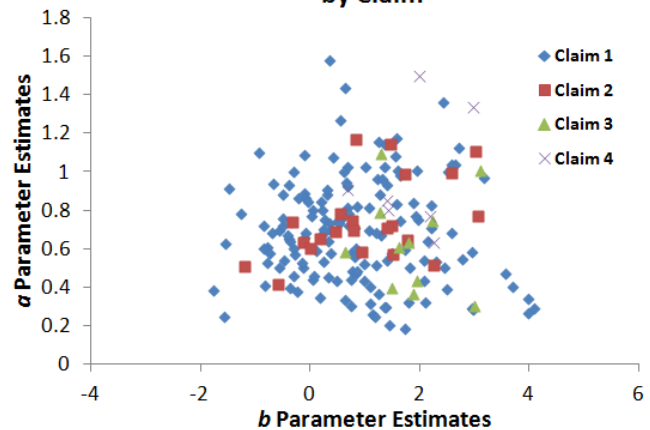
MATH G07 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
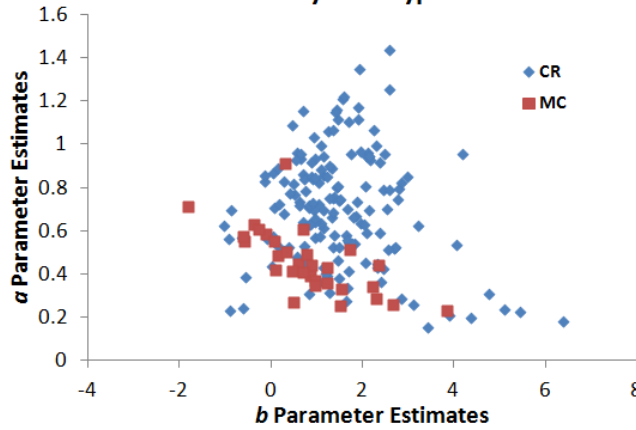
MATH G07 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
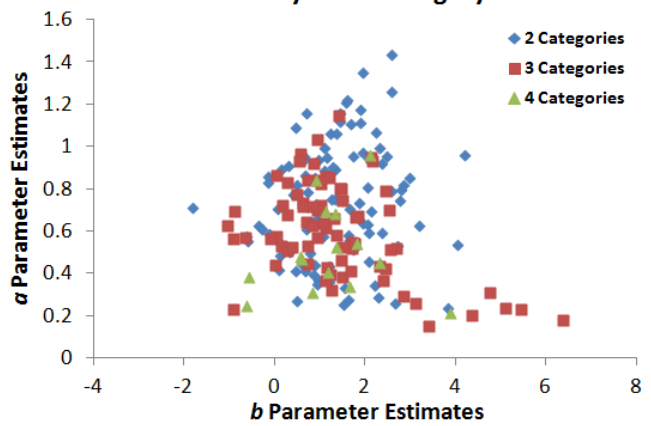
MATH G07 Scatter Plot of 2PL/GPC *a* and *b* by Claim
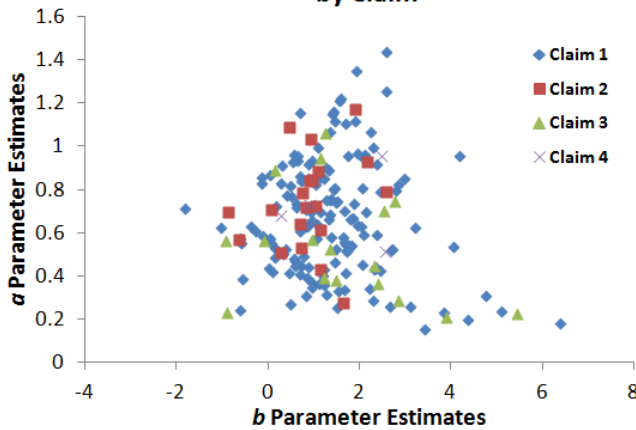
MATH G08 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
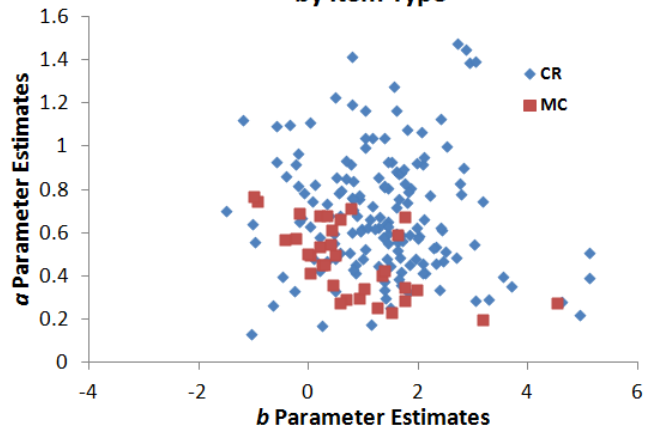
MATH G08 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
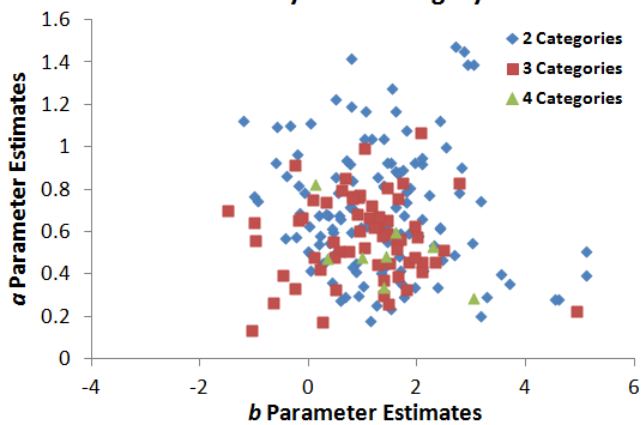
MATH G08 Scatter Plot of 2PL/GPC *a* and *b* by Claim
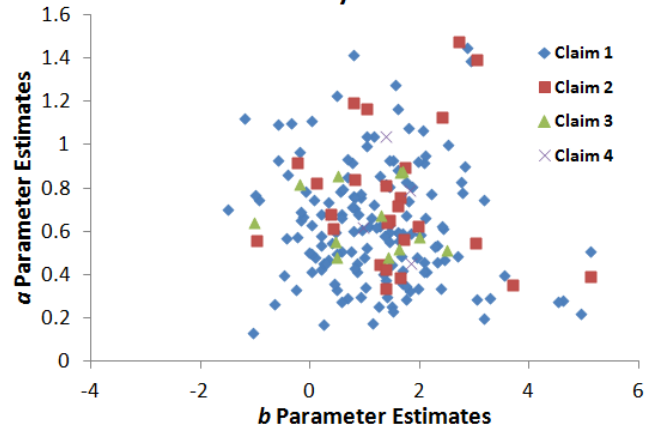
MATH G09 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
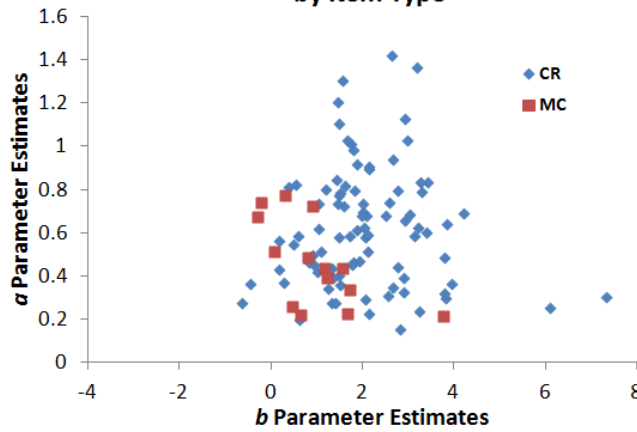
MATH G09 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
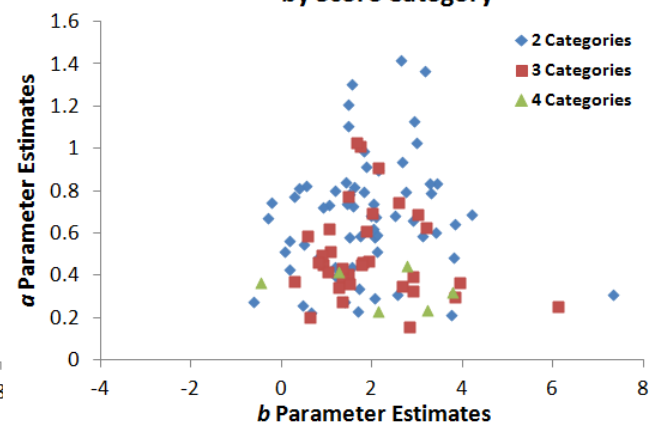
MATH G09 Scatter Plot of 2PL/GPC *a* and *b* by Claim
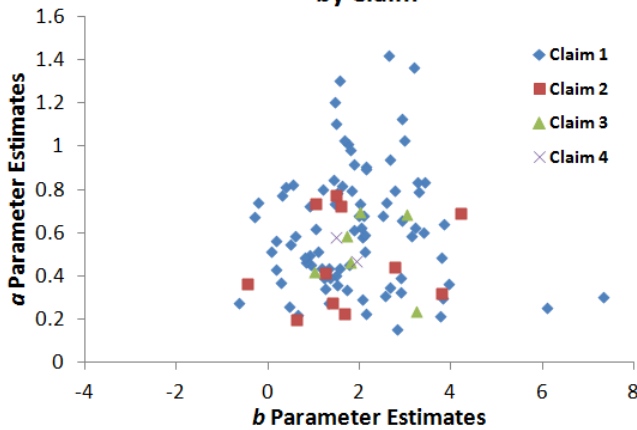
MATH G10 Scatter Plot of 2PL/GPC *a* and *b* by Item Type
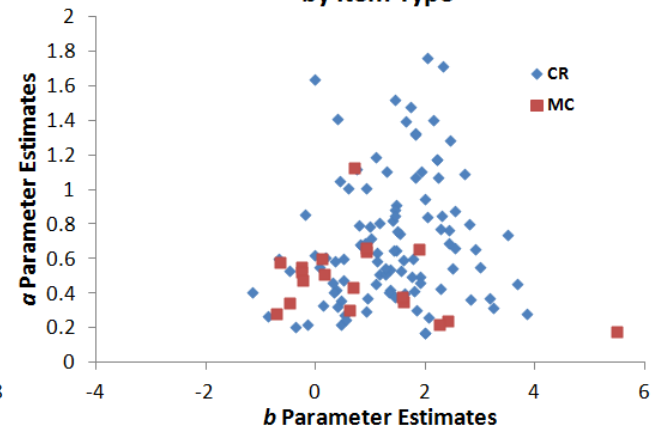
MATH G10 Scatter Plot of 2PL/GPC *a* and *b* by Score Category

MATH G10 Scatter Plot of 2PL/GPC *a* and *b* by Claim

MATH G11 Scatter Plot of 2PL/GPC *a* and *b* by Item Type

MATH G11 Scatter Plot of 2PL/GPC *a* and *b* by Score Category
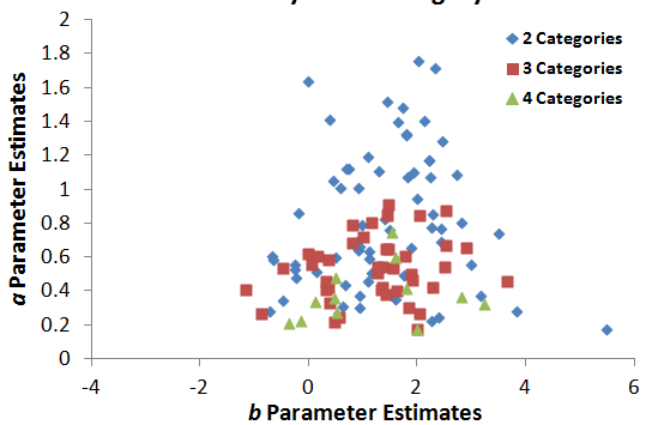
MATH G11 Scatter Plot of 2PL/GPC *a* and *b* by Claim
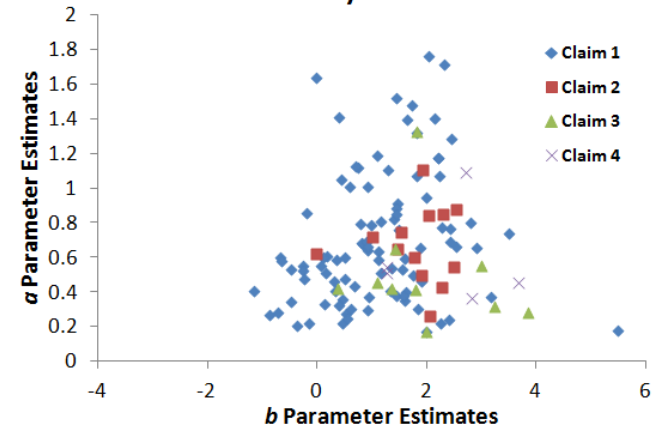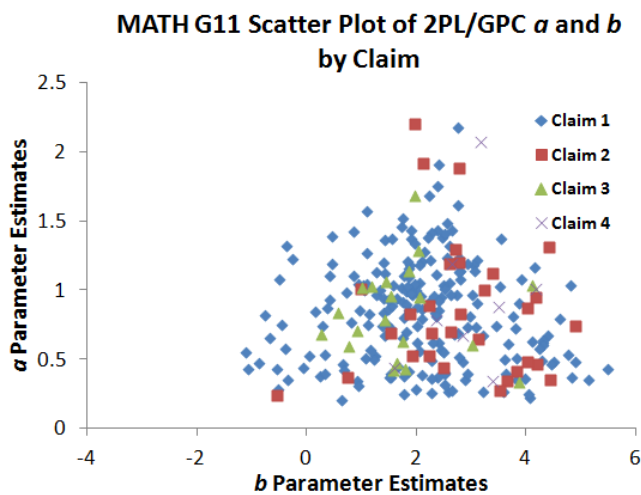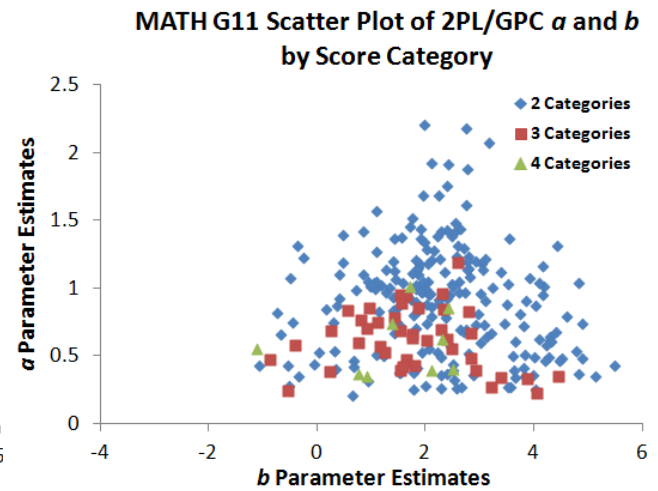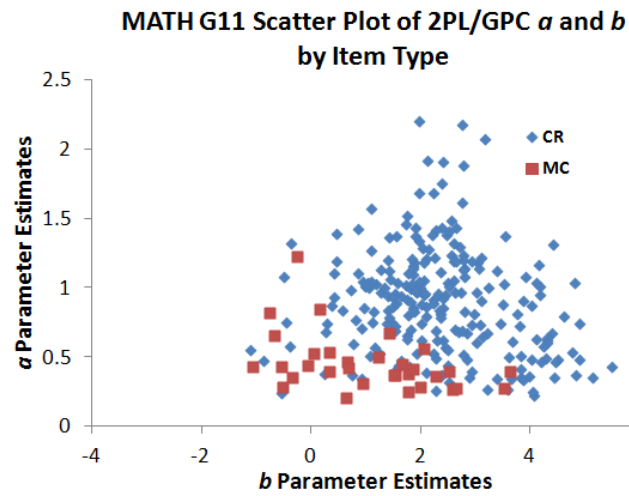
Figure B.3. ELA Scatter Plots of Theta Estimates across Different Model Combinations

Scatter Plot of Theta Estimates
ELA G05 (1PL/PC vs 2PL GPC)



Scatter Plot of Theta Estimates
ELA G05 (1PL/PC vs 3PL GPC)



Scatter Plot of Theta Estimates
ELA G05 (2PL/GPC vs 3PL GPC)



Scatter Plot of Theta Estimates
ELA G06 (1PL/PC vs 2PL GPC)



Scatter Plot of Theta Estimates
ELA G06 (1PL/PC vs 3PL GPC)



Scatter Plot of Theta Estimates
ELA G06 (2PL/GPC vs 3PL GPC)
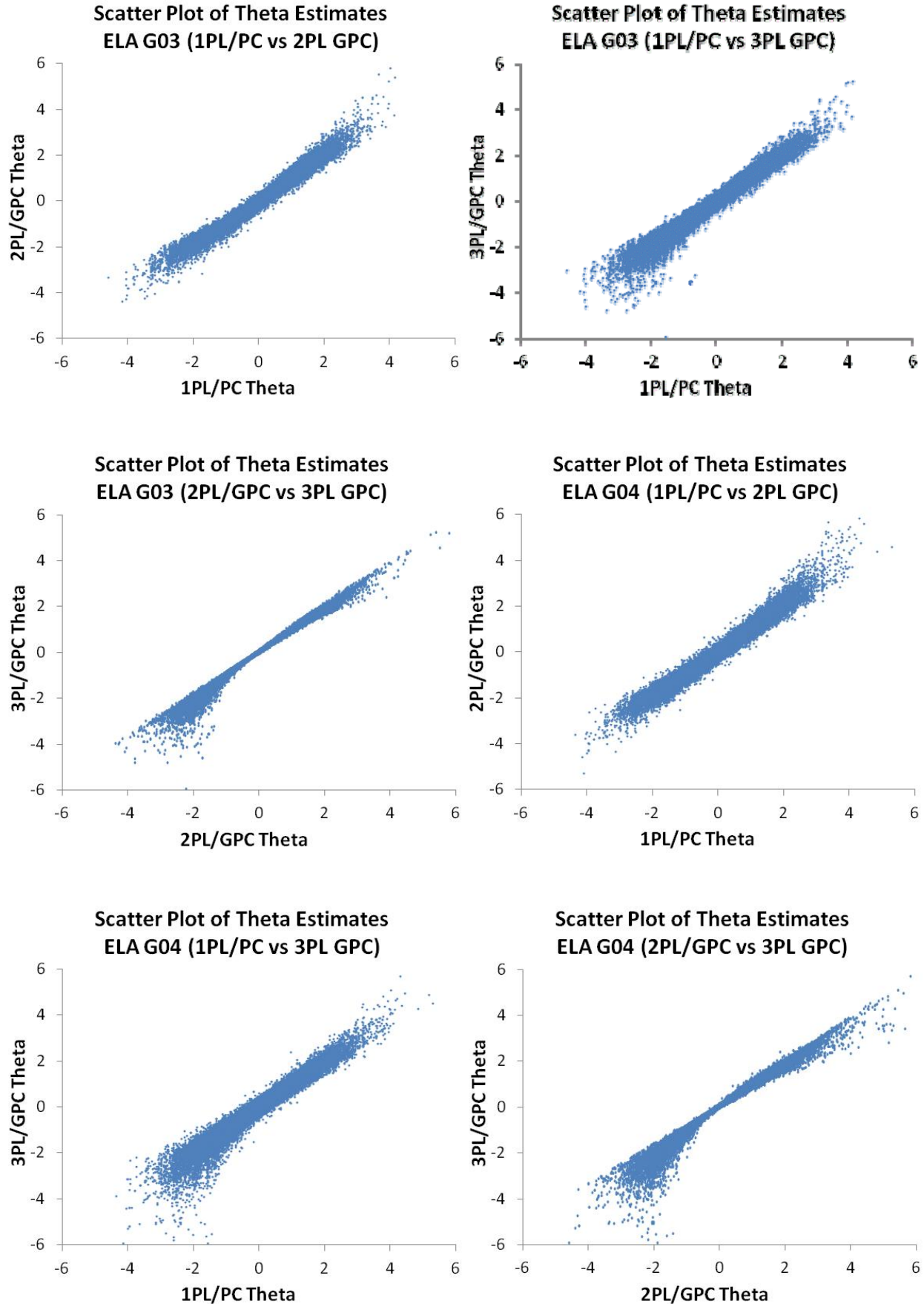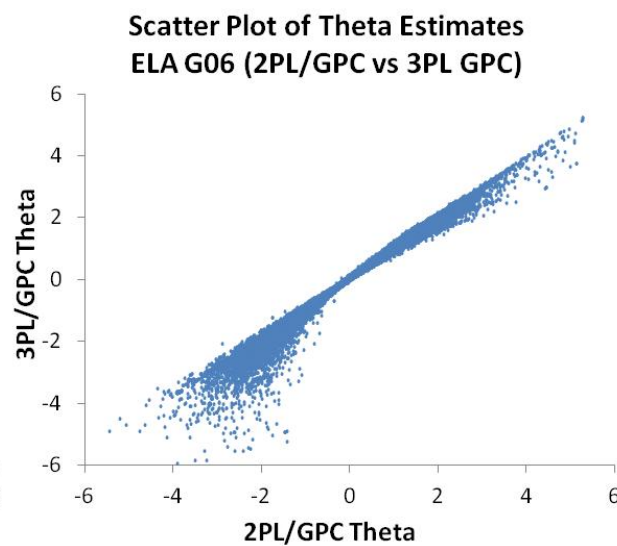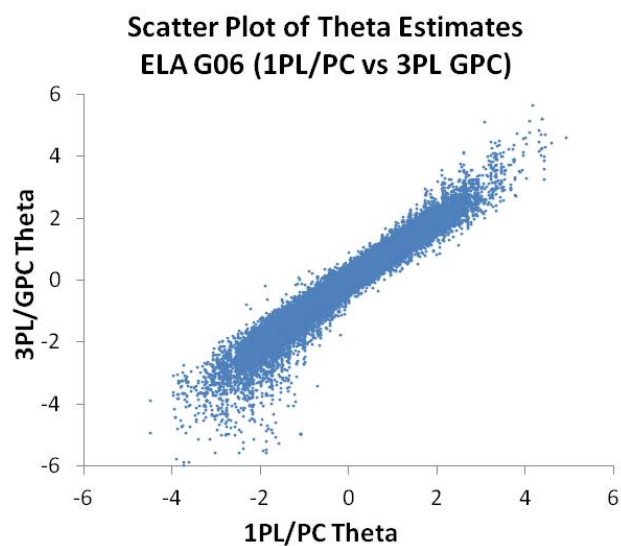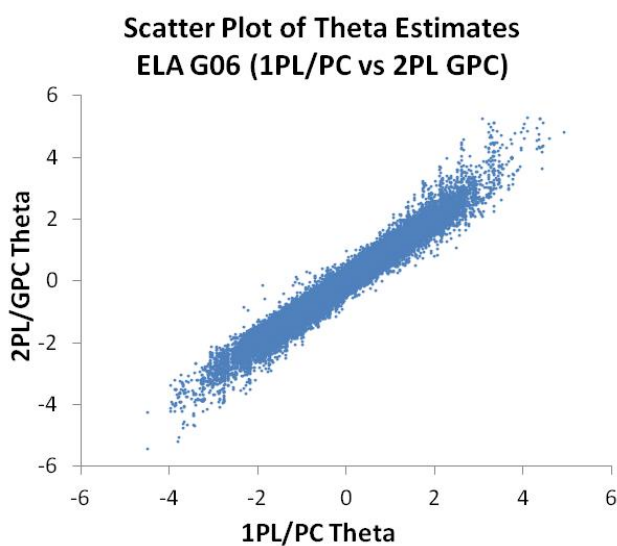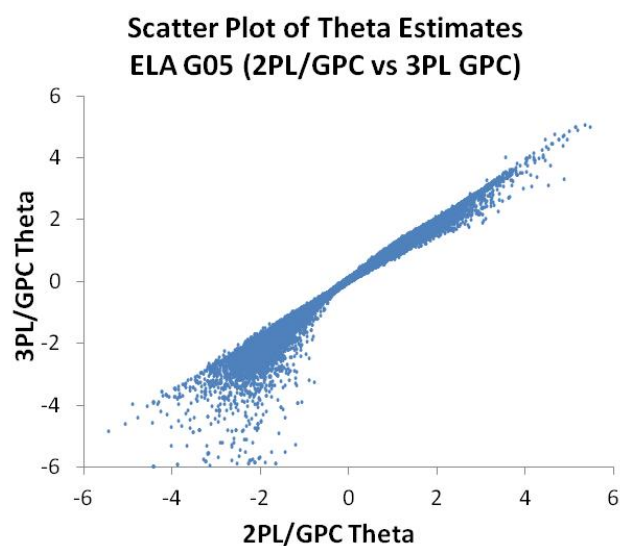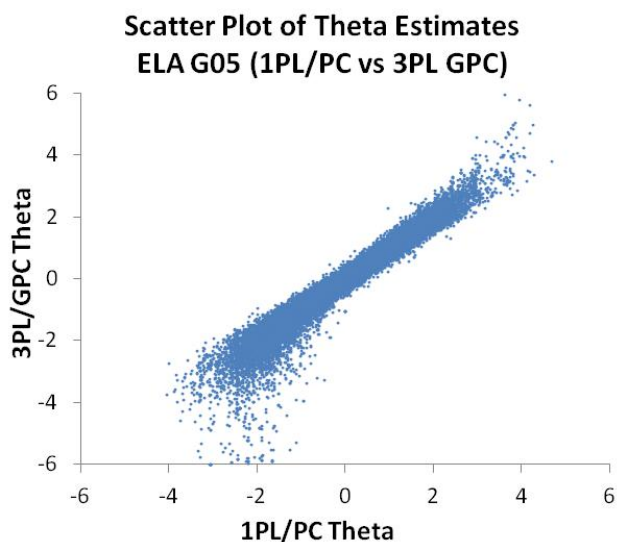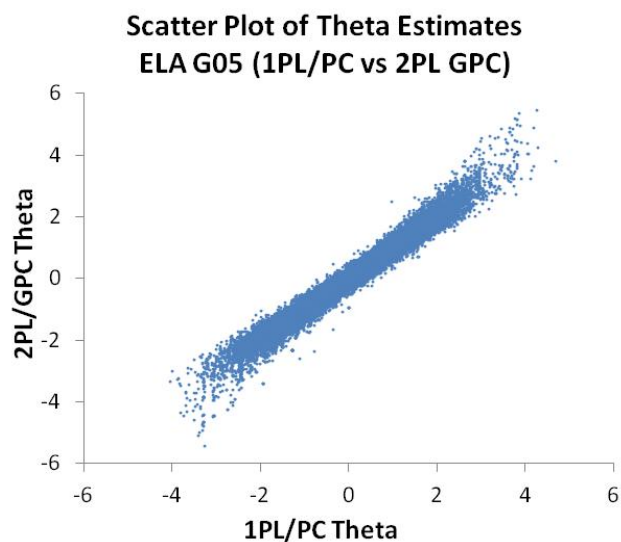
Scatter Plot of Theta Estimates
ELA G07 (1PL/PC vs 2PL GPC)
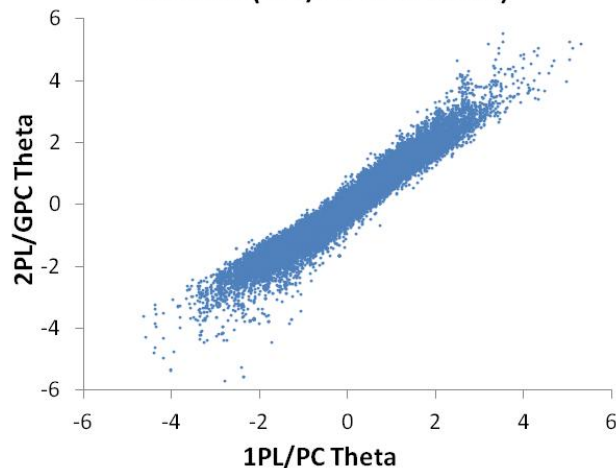
Scatter Plot of Theta Estimates
ELA G07 (1PL/PC vs 3PL GPC)
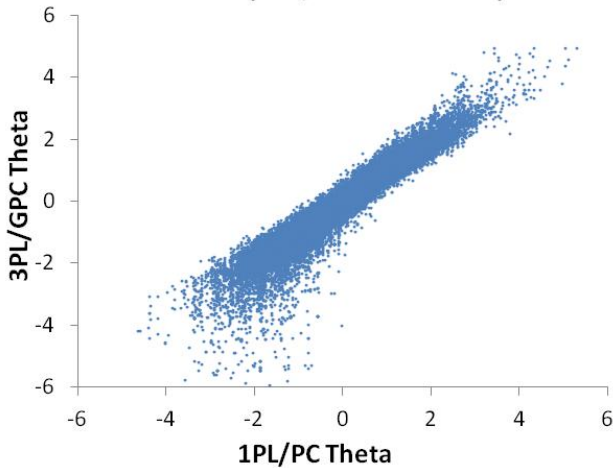
Scatter Plot of Theta Estimates
ELA G07 (2PL/GPC vs 3PL GPC)
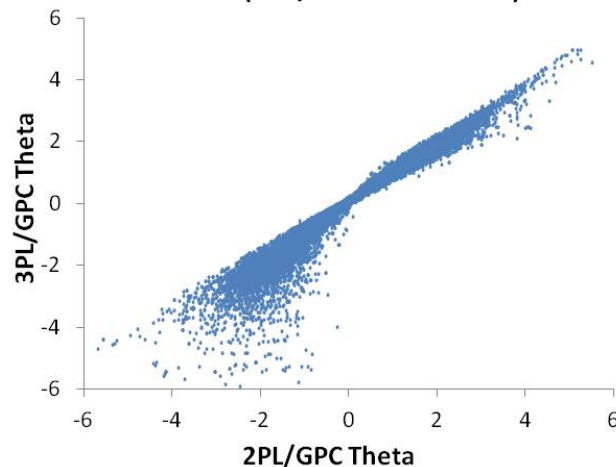
Scatter Plot of Theta Estimates
ELA G08 (1PL/PC vs 2PL GPC)
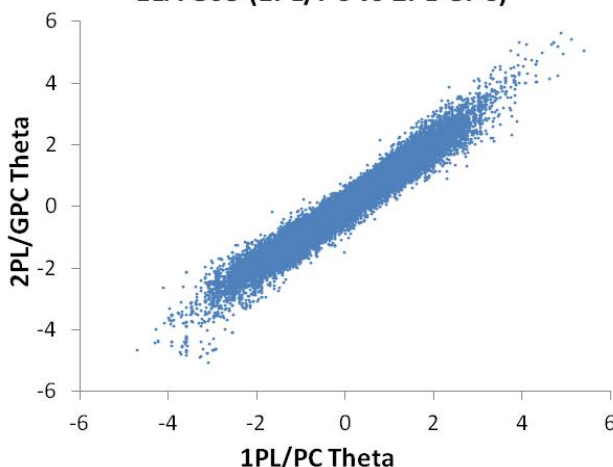
Scatter Plot of Theta Estimates
ELA G08 (1PL/PC vs 3PL GPC)
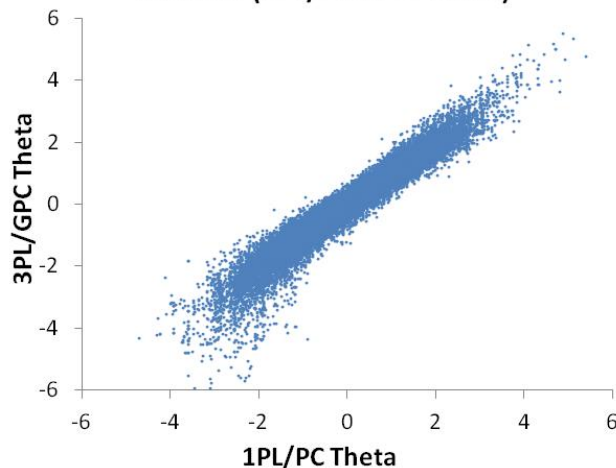
Scatter Plot of Theta Estimates
ELA G08 (2PL/GPC vs 3PL GPC)

Scatter Plot of Theta Estimates
ELA G09 (1PL/PC vs 2PL GPC)

Scatter Plot of Theta Estimates
ELA G09 (1PL/PC vs 3PL GPC)

Scatter Plot of Theta Estimates
ELA G09 (2PL/GPC vs 3PL GPC)

Scatter Plot of Theta Estimates
ELA G10 (1PL/PC vs 2PL GPC)

Scatter Plot of Theta Estimates
ELA G10 (1PL/PC vs 3PL GPC)

Scatter Plot of Theta Estimates
ELA G10 (2PL/GPC vs 3PL GPC)
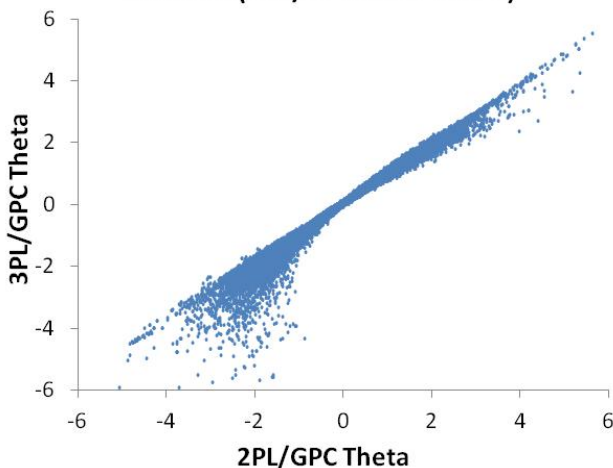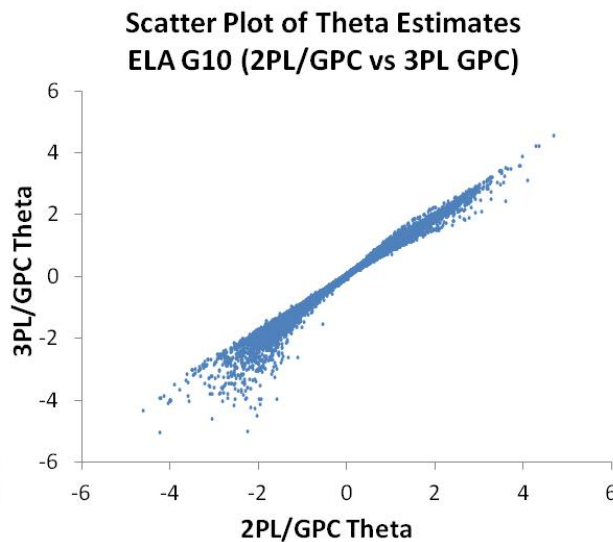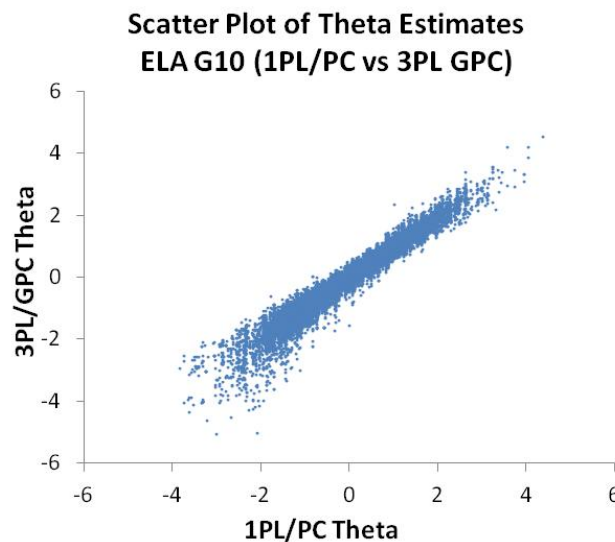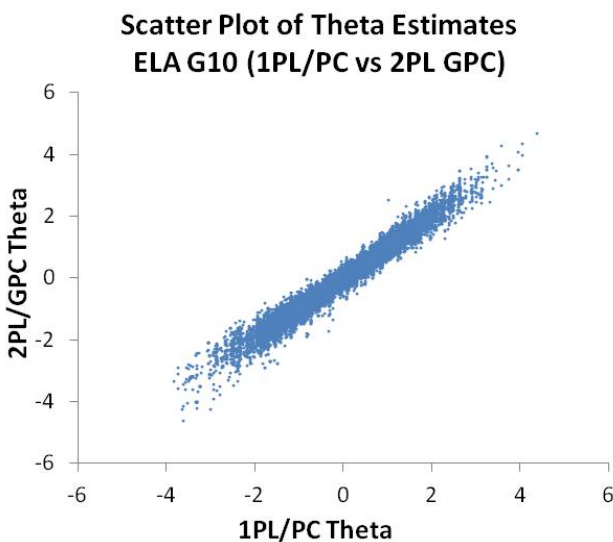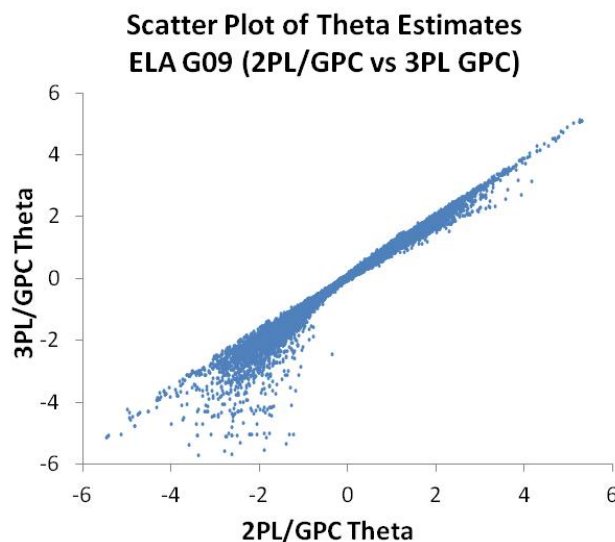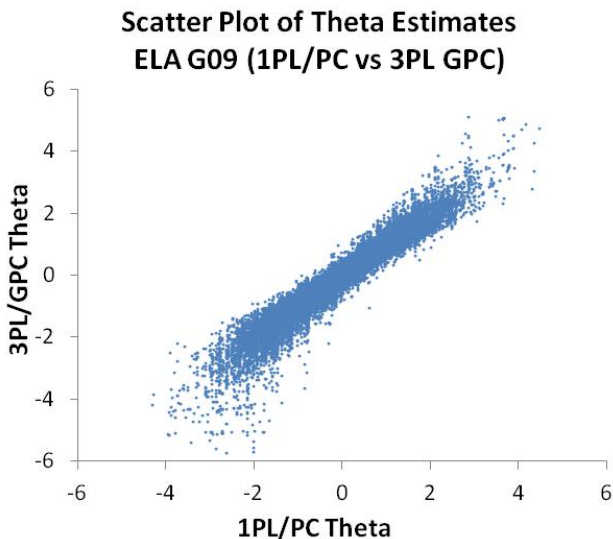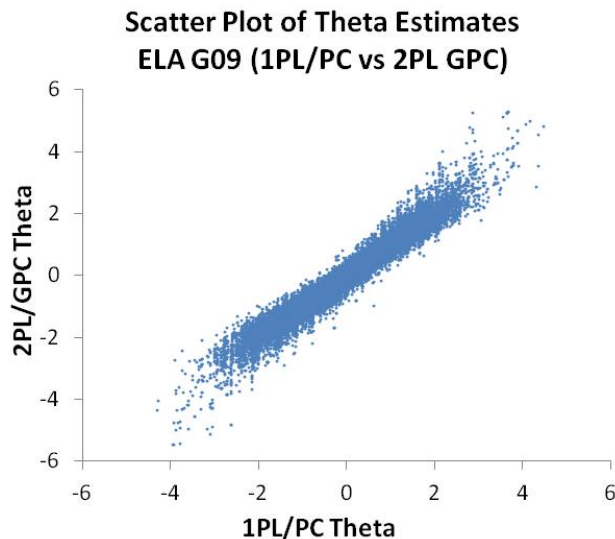
**Scatter Plot of Theta Estimates**
**ELA G11 (1PL/PC vs 2PL GPC)**
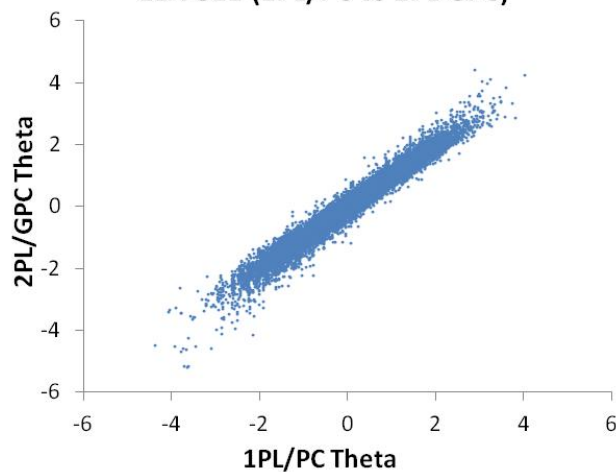
**Scatter Plot of Theta Estimates**
**ELA G11 (1PL/PC vs 3PL GPC)**
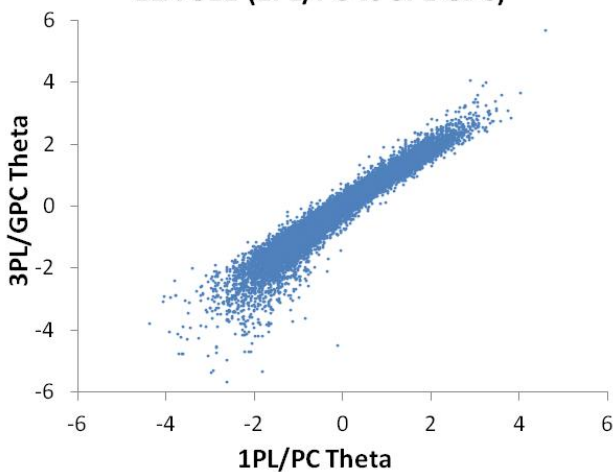
**Scatter Plot of Theta Estimates**
**ELA G11 (2PL/GPC vs 3PL GPC)**
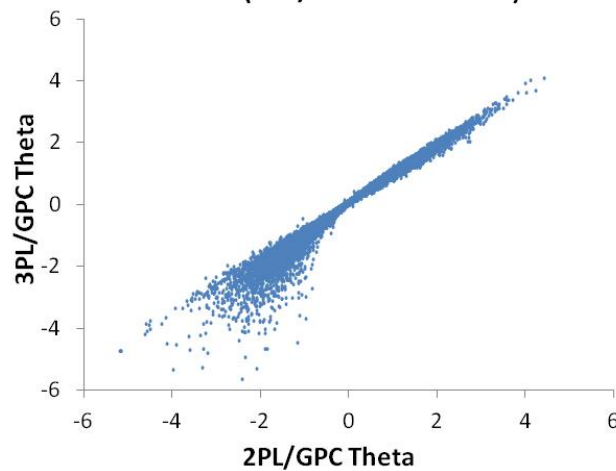
Figure B.4. Math Scatter Plots of Theta Estimates across Different Model Combinations

**Scatter Plot of Theta Estimates**
**Math G05 (1PL/PC vs 2PL GPC)**



**Scatter Plot of Theta Estimates**
**Math G05 (1PL/PC vs 3PL GPC)**



**Scatter Plot of Theta Estimates**
**Math G05 (2PL/GPC vs 3PL GPC)**



**Scatter Plot of Theta Estimates**
**Math G06 (1PL/PC vs 2PL GPC)**



**Scatter Plot of Theta Estimates**
**Math G06 (1PL/PC vs 3PL GPC)**



**Scatter Plot of Theta Estimates**
**Math G06 (2PL/GPC vs 3PL GPC)**

**Scatter Plot of Theta Estimates
Math G07 (1PL/PC vs 2PL GPC)**

**Scatter Plot of Theta Estimates
Math G07 (1PL/PC vs 3PL GPC)**

**Scatter Plot of Theta Estimates
Math G07 (2PL/GPC vs 3PL GPC)**

**Scatter Plot of Theta Estimates
Math G08 (1PL/PC vs 2PL GPC)**

**Scatter Plot of Theta Estimates
Math G08 (1PL/PC vs 3PL GPC)**

**Scatter Plot of Theta Estimates
Math G08 (2PL/GPC vs 3PL GPC)**

Scatter Plot of Theta Estimates
Math G09 (1PL/PC vs 2PL GPC)

Scatter Plot of Theta Estimates
Math G09 (1PL/PC vs 3PL GPC)

Scatter Plot of Theta Estimates
Math G09 (2PL/GPC vs 3PL GPC)

Scatter Plot of Theta Estimates
Math G10 (1PL/PC vs 2PL GPC)

Scatter Plot of Theta Estimates
Math G10 (1PL/PC vs 3PL GPC)

Scatter Plot of Theta Estimates
Math G10 (2PL/GPC vs 3PL GPC)

**Scatter Plot of Theta Estimates Math G11 (1PL/PC vs 2PL GPC)**

**Scatter Plot of Theta Estimates Math G11 (1PL/PC vs 3PL GPC)**
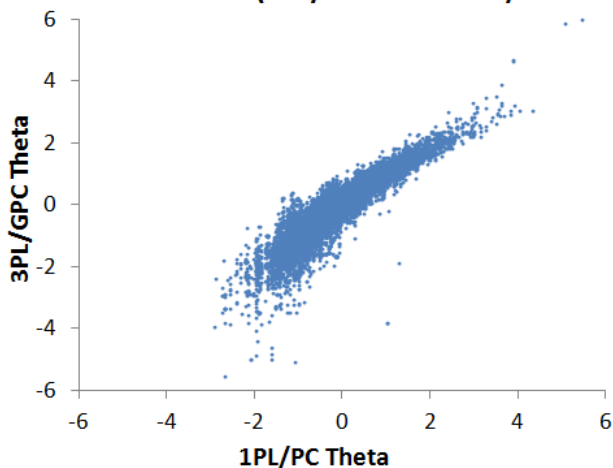
**Scatter Plot of Theta Estimates Math G11 (2PL/GPC vs 3PL GPC)**
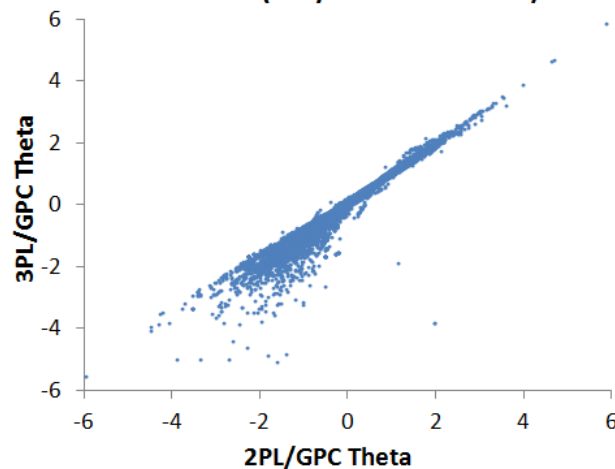
Table B.1. ELA Items Receiving Pre-treatment before Calibration based on Data Clearning Procedure

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 3 | 3 | 52402 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 3 | 3 | 52411 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 3 | 3 | 53779 | CAT | 2 | Dropped as suggested by content review |
| 3 | 3 | 53801 | CAT | 2 | Dropped due to low item-total correlation |
| 3 | 3 | 53925 | CAT | 3 | Dropped as suggested by content review |
| 3 | 3 | 54099 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 54163 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 54219 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 54223 | CAT | 4 | Dropped due to low item-total correlation |
| 3 | 3 | 54253 | CAT | 2 | Dropped due to low item-total correlation |
| 3 | 3 | 54303 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 54319 | CAT | 3 | Dropped due to low item-total correlation |
| 3 | 3 | 56130A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56130B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56133 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 3 | 3 | 56134A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 3 | 3 | 56134B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 3 | 3 | 56194A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56194B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56199A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56199B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 3 | 3 | 56325A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 3 | 3 | 56325B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 3 | 4 | 54616 | CAT | 4 | Dropped due to low item-total correlation |
| 3 | 4 | 54982 | CAT | 3 | Dropped as suggested by content review |
| 3 | 4 | 56186 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56187 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56188A | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56188B | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56188C | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56244 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 3 | 4 | 56245 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56247 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56248A | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56248B | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56248C | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56258 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56259 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56261 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56263A | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56263B | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56263C | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56299 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56302 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56309A | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56309B | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56309C | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56311 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56312 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56313A | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56313B | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56313C | PT | 2 | Dropped due to no scored responses |
| 3 | 4 | 56461 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56462 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56463 | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56464A | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56464B | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56464C | PT | 4 | Dropped due to no scored responses |
| 3 | 4 | 56468 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 52402 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 4 | 3 | 53925 | CAT | 3 | Dropped as suggested by content review |
| 4 | 3 | 54099 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 3 | 54253 | CAT | 2 | Dropped due to low item-total correlation |
| 4 | 3 | 54303 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 3 | 56126 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56128 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56130A | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56130B | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56130C | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56133 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56134A | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56134B | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56134C | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56189 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 4 | 3 | 56192 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56194A | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56194B | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56194C | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56197 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56198 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56199A | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56199B | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56199C | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56324 | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56325A | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56325B | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56325C | PT | 2 | Dropped due to no scored responses |
| 4 | 3 | 56390 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56410 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56411 | PT | 4 | Dropped due to no scored responses |
| 4 | 3 | 56467 | PT | 4 | Dropped due to no scored responses |
| 4 | 4 | 54490 | CAT | 2 | Dropped due to low item-total correlation |
| 4 | 4 | 54500 | CAT | 2 | Dropped as suggested by content review |
| 4 | 4 | 54540 | CAT | 2 | Dropped as suggested by content review |
| 4 | 4 | 54568 | CAT | 2 | Dropped as suggested by content review |
| 4 | 4 | 54580 | CAT | 2 | Dropped as suggested by content review |
| 4 | 4 | 54588 | CAT | 2 | Dropped as suggested by content review |
| 4 | 4 | 54616 | CAT | 4 | Dropped due to low item-total correlation |
| 4 | 4 | 54634 | CAT | 4 | Dropped due to low item-total correlation |
| 4 | 4 | 54982 | CAT | 3 | Dropped as suggested by content review |
| 4 | 4 | 55023 | CAT | 3 | Dropped due to low item-total correlation |
| 4 | 4 | 55025 | CAT | 3 | Dropped due to low item-total correlation |
| 4 | 4 | 55027 | CAT | 3 | Dropped due to low item-total correlation |
| 4 | 4 | 55350 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 55368 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 55444 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 55667 | CAT | 3 | Dropped as suggested by content review |
| 4 | 4 | 55688 | CAT | 3 | Dropped due to low item-total correlation |
| 4 | 4 | 55738 | CAT | 3 | Dropped as suggested by content review |
| 4 | 4 | 55742 | CAT | 3 | Dropped as suggested by content review |
| 4 | 4 | 56188A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56188B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56248A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56248B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 4 | 4 | 56263A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56263B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56309A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56309B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56313A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56313B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 4 | 4 | 56462 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 4 | 4 | 56464A | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 4 | 4 | 56464B | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 4 | 5 | 54674 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 5 | 54676 | CAT | 1 | Dropped as suggested by content review |
| 4 | 5 | 55099 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55105 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55109 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55110A | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 55110B | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 55110C | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 55542 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55544 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55545 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 55547A | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 55547B | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 55547C | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56191 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56193 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56195 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56196A | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56196B | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56196C | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56271 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56272 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56273 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56274A | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56274B | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56274C | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56320 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 4 | 5 | 56321 | PT | 4 | Dropped due to no scored responses |
| 4 | 5 | 56322A | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56322B | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56322C | PT | 2 | Dropped due to no scored responses |
| 4 | 5 | 56469 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 54490 | CAT | 2 | Dropped due to low item-total correlation |
| 5 | 4 | 54568 | CAT | 2 | Dropped as suggested by content review |
| 5 | 4 | 55667 | CAT | 3 | Dropped as suggested by content review |
| 5 | 4 | 55738 | CAT | 3 | Dropped as suggested by content review |
| 5 | 4 | 56186 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56187 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56188A | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56188B | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56188C | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56244 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56245 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56247 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56248A | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56248B | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56248C | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56258 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56259 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56261 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56263A | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56263B | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56263C | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56299 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56302 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56309A | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56309B | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56309C | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56311 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56312 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56313A | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56313B | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56313C | PT | 2 | Dropped due to no scored responses |
| 5 | 4 | 56461 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56462 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56463 | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56464A | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56464B | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56464C | PT | 4 | Dropped due to no scored responses |
| 5 | 4 | 56468 | PT | 4 | Dropped due to no scored responses |
| 5 | 5 | 52315 | CAT | 3 | Dropped due to low item-total correlation |

99

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 5 | 5 | 53652 | CAT | 1 | Dropped as suggested by content review |
| 5 | 5 | 53663 | CAT | 1 | Dropped due to low item-total correlation |
| 5 | 5 | 54674 | CAT | 1 | Dropped due to low item-total correlation |
| 5 | 5 | 54676 | CAT | 1 | Dropped as suggested by content review |
| 5 | 5 | 54764 | CAT | 1 | Dropped as suggested by content review |
| 5 | 5 | 54858 | CAT | 2 | Dropped due to low item-total correlation |
| 5 | 5 | 54922 | CAT | 3 | Dropped due to low item-total correlation |
| 5 | 5 | 54940 | CAT | 4 | Dropped due to low item-total correlation |
| 5 | 5 | 55110A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 55110B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 5 | 5 | 55547A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 55547B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 56196A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 56196B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 56274A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 56274B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 5 | 5 | 56322A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 5 | 56322B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 5 | 6 | 50577 | CAT | 2 | Dropped due to no scored responses |
| 5 | 6 | 52257 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 52268 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 52269 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 52390 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 52398A | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 52398B | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 52398C | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 52645 | CAT | 1 | Dropped as suggested by content review |
| 5 | 6 | 52647 | CAT | 1 | Dropped as suggested by content review |
| 5 | 6 | 52689 | CAT | 1 | Dropped due to low item-total correlation |
| 5 | 6 | 52801 | CAT | 2 | Dropped as suggested by content review |
| 5 | 6 | 52849 | CAT | 3 | Dropped as suggested by content review |
| 5 | 6 | 52855 | CAT | 3 | Dropped as suggested by content review |
| 5 | 6 | 53021 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 53022 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 5 | 6 | 53023 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 53024A | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 53024B | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 53024C | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55085 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55086 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55088A | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55088B | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55088C | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55089 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55090 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55092 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55093 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55094A | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55094B | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55094C | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55095 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55098 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55103A | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55103B | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55103C | PT | 2 | Dropped due to no scored responses |
| 5 | 6 | 55631 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55922 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55923 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55925 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55926 | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55927A | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55927B | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 55927C | PT | 4 | Dropped due to no scored responses |
| 5 | 6 | 56012 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 5 | 6 | 56121 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 6 | 5 | 52315 | CAT | 3 | Dropped due to low item-total correlation |
| 6 | 5 | 53652 | CAT | 1 | Dropped as suggested by content review |
| 6 | 5 | 53663 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 5 | 55099 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55105 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55109 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55110A | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 55110B | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 55110C | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 55542 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55544 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55545 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 55547A | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 6 | 5 | 55547B | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 55547C | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56191 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56193 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56195 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56196A | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56196B | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56196C | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56252 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56271 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56272 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56273 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56274A | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56274B | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56274C | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56320 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56321 | PT | 4 | Dropped due to no scored responses |
| 6 | 5 | 56322A | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56322B | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56322C | PT | 2 | Dropped due to no scored responses |
| 6 | 5 | 56469 | PT | 4 | Dropped due to no scored responses |
| 6 | 6 | 46544 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 6 | 6 | 47824 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 47844 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 48230 | CAT | 4 | Dropped as suggested by content review |
| 6 | 6 | 48333 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 6 | 6 | 48350 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 48701 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 48799 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 48801 | CAT | 3 | Dropped due to low item-total correlation |
| 6 | 6 | 52398A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 52398B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 52645 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 52647 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 52673 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 52689 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 52707 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 52712 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 52716 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 52718 | CAT | 1 | Dropped due to low item-total correlation |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 6 | 6 | 52766 | CAT | 2 | Dropped due to low item-total correlation |
| 6 | 6 | 52783 | CAT | 2 | Dropped due to low item-total correlation |
| 6 | 6 | 52791 | CAT | 2 | Dropped due to low item-total correlation |
| 6 | 6 | 52801 | CAT | 2 | Dropped as suggested by content review |
| 6 | 6 | 52825 | CAT | 2 | Dropped due to low item-total correlation |
| 6 | 6 | 52849 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 52855 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 52859 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 52873 | CAT | 3 | Dropped due to low item-total correlation |
| 6 | 6 | 52895 | CAT | 4 | Dropped due to low item-total correlation |
| 6 | 6 | 53024A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 6 | 6 | 53024B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 6 | 6 | 55088A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 55088B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 55094A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 55094B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 6 | 6 | 55103A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 6 | 6 | 55103B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 6 | 6 | 55927A | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 55927B | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 6 | 6 | 56121 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 6 | 7 | 46442 | CAT | 3 | Dropped as suggested by content review |
| 6 | 7 | 46454 | CAT | 3 | Dropped due to low item-total correlation |
| 6 | 7 | 47493 | CAT | 1 | Dropped as suggested by content review |
| 6 | 7 | 47888 | CAT | 2 | Dropped as suggested by content review |
| 6 | 7 | 52478 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 52480 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 52587A | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 52587B | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 52587C | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 52780 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53018 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53019A | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53019B | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 6 | 7 | 53019C | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53025 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53026 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53027 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53028A | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53028B | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53028C | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53029 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53030 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53031 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53032A | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53032B | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53032C | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53126 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53127 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53128 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53129A | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53129B | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53129C | PT | 2 | Dropped due to no scored responses |
| 6 | 7 | 53768 | PT | 4 | Dropped due to no scored responses |
| 6 | 7 | 53769 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 46544 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 7 | 6 | 47824 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 47844 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 48230 | CAT | 4 | Dropped as suggested by content review |
| 7 | 6 | 48333 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 7 | 6 | 48350 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 48701 | CAT | 3 | Dropped as suggested by content review |
| 7 | 6 | 48799 | CAT | 3 | Dropped as suggested by content review |
| 7 | 6 | 48801 | CAT | 3 | Dropped due to low item-total correlation |
| 7 | 6 | 52257 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 52268 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 52269 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 52390 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 52398A | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 52398B | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 52398C | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 53021 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 53023 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 53024A | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 53024B | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 53024C | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 7 | 6 | 55085 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55086 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55088A | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55088B | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55088C | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55089 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55090 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55092 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55093 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55094A | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55094B | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55094C | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55095 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55098 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55103A | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55103B | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55103C | PT | 2 | Dropped due to no scored responses |
| 7 | 6 | 55920 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55922 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55923 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55925 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55926 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55927A | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55927B | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 55927C | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 56012 | PT | 4 | Dropped due to no scored responses |
| 7 | 6 | 56121 | PT | 4 | Dropped due to no scored responses |
| 7 | 7 | 46117 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 46264 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 46424 | CAT | 3 | Dropped as suggested by content review |
| 7 | 7 | 46442 | CAT | 3 | Dropped as suggested by content review |
| 7 | 7 | 46454 | CAT | 3 | Dropped due to low item-total correlation |
| 7 | 7 | 46472 | CAT | 3 | Dropped as suggested by content review |
| 7 | 7 | 47369 | CAT | 4 | Dropped as suggested by content review |
| 7 | 7 | 47471 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 47493 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 47509 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 47872 | CAT | 2 | Dropped due to low item-total correlation |
| 7 | 7 | 47882 | CAT | 2 | Dropped due to low item-total correlation |
| 7 | 7 | 47888 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 47928 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 48405 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 52587A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |

105

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 7 | 7 | 52587B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53019A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53019B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53028A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53028B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53032A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53032B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53129A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 7 | 53129B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 7 | 8 | 46223 | CAT | 2 | Dropped due to no scored responses |
| 7 | 8 | 46507 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 7 | 8 | 46509 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 7 | 8 | 47657 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 8 | 47689 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 8 | 47691 | CAT | 1 | Dropped as suggested by content review |
| 7 | 8 | 47799 | CAT | 3 | Dropped as suggested by content review |
| 7 | 8 | 48264 | CAT | 1 | Dropped as suggested by content review |
| 7 | 8 | 48309 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 7 | 8 | 48344 | CAT | 3 | Dropped due to low item-total correlation |
| 7 | 8 | 52467 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 52472 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 52473 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 52477 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 52586A | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 52586B | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 52586C | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53038 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53039 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53040 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53041A | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53041B | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53041C | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53042 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 7 | 8 | 53043 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53044 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53045A | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53045B | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53045C | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53046 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53047 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53048 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53049 | PT | 4 | Dropped due to low item-total correlation |
| 7 | 8 | 53050A | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53050B | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53050C | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53130 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53131 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53132 | PT | 4 | Dropped due to no scored responses |
| 7 | 8 | 53133A | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53133B | PT | 2 | Dropped due to no scored responses |
| 7 | 8 | 53133C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 46095 | CAT | 2 | Dropped due to no scored responses |
| 8 | 7 | 46424 | CAT | 3 | Dropped as suggested by content review |
| 8 | 7 | 46472 | CAT | 3 | Dropped as suggested by content review |
| 8 | 7 | 47369 | CAT | 4 | Dropped as suggested by content review |
| 8 | 7 | 47471 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 7 | 52478 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 52480 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 52587A | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 52587B | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 52587C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 52780 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53018 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53019A | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53019B | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53019C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53025 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53026 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53027 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53028A | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53028B | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53028C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53029 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53030 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53031 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53032A | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53032B | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 8 | 7 | 53032C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53126 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53127 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53128 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53129A | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53129B | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53129C | PT | 2 | Dropped due to no scored responses |
| 8 | 7 | 53768 | PT | 4 | Dropped due to no scored responses |
| 8 | 7 | 53769 | PT | 4 | Dropped due to no scored responses |
| 8 | 8 | 46223 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 46225 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 46507 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 47243 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47275 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 8 | 47283 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 8 | 47311 | CAT | 3 | Dropped as suggested by content review |
| 8 | 8 | 47317 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 8 | 47323 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 8 | 47427 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 47591 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47599 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47603 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47627 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47647 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47649 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 47657 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 47667 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47681 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 47689 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 47691 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47695 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47735 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47741 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 47799 | CAT | 3 | Dropped as suggested by content review |
| 8 | 8 | 47948 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 47952 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 8 | 47974 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 8 | 47994 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 48010 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 8 | 48036 | CAT | 2 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 8 | 8 | 48186 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 48264 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 48309 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 48335 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 8 | 8 | 48344 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 8 | 52586A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 52586B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53041A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53041B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53045A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 8 | 8 | 53045B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 8 | 8 | 53050A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53050B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53133A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 8 | 53133B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 8 | 9 | 46724 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 46726 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 46728 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 47779 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 47787 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 47789 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 48055 | CAT | 2 | Dropped as suggested by content review |
| 8 | 9 | 48067 | CAT | 2 | Dropped as suggested by content review |
| 8 | 9 | 48259 | CAT | 1 | Dropped as suggested by content review |
| 8 | 9 | 48607 | CAT | 3 | Dropped due to low item-total correlation |
| 8 | 9 | 53033 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53034 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53035 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53036 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53037A | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 53037B | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 53037C | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 53058 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 8 | 9 | 53059 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53060 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 53061A | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 53061B | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 53061C | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55091 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55096 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55102 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55108A | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55108B | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55108C | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55111 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55112 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55113 | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55114A | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55114B | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55114C | PT | 1 | Dropped due to no scored responses |
| 8 | 9 | 55553 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55556 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55557 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55559A | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55559B | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55559C | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55598 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55600 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55601 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55624 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55625 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55626 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55627A | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55627B | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55627C | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55902 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55903 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55904 | PT | 4 | Dropped due to no scored responses |
| 8 | 9 | 55905A | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55905B | PT | 2 | Dropped due to no scored responses |
| 8 | 9 | 55905C | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 47275 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 8 | 47311 | CAT | 3 | Dropped as suggested by content review |
| 9 | 8 | 47317 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 8 | 47427 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 9 | 8 | 47627 | CAT | 1 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 9 | 8 | 47667 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 47695 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 47948 | CAT | 2 | Dropped as suggested by content review |
| 9 | 8 | 47994 | CAT | 2 | Dropped as suggested by content review |
| 9 | 8 | 48186 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 8 | 52467 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 52472 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 52473 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 52477 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 52586A | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 52586B | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 52586C | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53038 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53039 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53040 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53041A | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53041B | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53041C | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53042 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53043 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53044 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53045A | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53045B | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53045C | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53046 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53047 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53048 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53050A | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53050B | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53050C | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53130 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53131 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53132 | PT | 4 | Dropped due to no scored responses |
| 9 | 8 | 53133A | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53133B | PT | 2 | Dropped due to no scored responses |
| 9 | 8 | 53133C | PT | 2 | Dropped due to no scored responses |
| 9 | 9 | 46724 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 46726 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 46728 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 47779 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 47787 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 47789 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 48055 | CAT | 2 | Dropped as suggested by content review |
| 9 | 9 | 48067 | CAT | 2 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 9 | 9 | 48259 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 48607 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 53037A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 53037B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 53060 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 9 | 53061A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 53061B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 53390 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 53392 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 53421 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 53435 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 53439 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 53473 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 53488 | CAT | 3 | Dropped as suggested by content review |
| 9 | 9 | 53490 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 53492 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 53630 | CAT | 2 | Dropped due to low item-total correlation |
| 9 | 9 | 55108A | PT | 1 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 55108B | PT | 1 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 55112 | PT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 9 | 55114A | PT | 1 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 55114B | PT | 1 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 55559A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 9 | 9 | 55559B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 55627A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 55627B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 55905A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 9 | 55905B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 9 | 10 | 51542 | CAT | 2 | Dropped due to no scored responses |
| 9 | 10 | 51554 | CAT | 2 | Dropped due to no scored responses |
| 9 | 10 | 53530 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 9 | 10 | 53538 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 10 | 53548 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 10 | 53558 | CAT | 1 | Dropped as suggested by content review |
| 9 | 10 | 53594 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 53596 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 53598 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 53600 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 53606 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 10 | 53612 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 10 | 53620 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 10 | 53624 | CAT | 4 | Dropped as suggested by content review |
| 9 | 10 | 55097 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55101 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55104 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55107A | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55107B | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55107C | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55258 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55259 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55260A | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55260B | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55260C | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55619 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55620 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55621 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55622 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55623A | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55623B | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55623C | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55918 | PT | 4 | Dropped due to low item-total correlation |
| 9 | 10 | 55930 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55931 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55932 | PT | 4 | Dropped due to no scored responses |
| 9 | 10 | 55933A | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55933B | PT | 2 | Dropped due to no scored responses |
| 9 | 10 | 55933C | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53033 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53034 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53035 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53036 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53037A | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53037B | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53037C | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53058 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 10 | 9 | 53059 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53060 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 53061A | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53061B | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53061C | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 53378 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 53390 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 53392 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 53419 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 53421 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 53435 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 53439 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 53473 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 9 | 53488 | CAT | 3 | Dropped as suggested by content review |
| 10 | 9 | 53490 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 9 | 53492 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 9 | 53630 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 9 | 55091 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55096 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55102 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55108A | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55108B | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55108C | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55111 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55112 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55113 | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55114A | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55114B | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55114C | PT | 1 | Dropped due to no scored responses |
| 10 | 9 | 55553 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55556 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55557 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55559A | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55559B | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55559C | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55598 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55600 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55601 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55624 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55625 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55626 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55627A | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55627B | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55627C | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 10 | 9 | 55902 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55903 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55904 | PT | 4 | Dropped due to no scored responses |
| 10 | 9 | 55905A | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55905B | PT | 2 | Dropped due to no scored responses |
| 10 | 9 | 55905C | PT | 2 | Dropped due to no scored responses |
| 10 | 10 | 48704 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 48719 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 48726 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 48846 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 48897 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 48907 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 48909 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 49356 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 49530 | CAT | 4 | Dropped due to low item-total correlation |
| 10 | 10 | 49532 | CAT | 4 | Dropped due to low item-total correlation |
| 10 | 10 | 49536 | CAT | 4 | Dropped as suggested by content review |
| 10 | 10 | 49599 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 49603 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 53538 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 53558 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 53594 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 53596 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 53598 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 53600 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 53606 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 53612 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 53620 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 10 | 53624 | CAT | 4 | Dropped as suggested by content review |
| 10 | 10 | 55107A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55107B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55260A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55260B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55623A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55623B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 10 | 10 | 55918 | PT | 4 | Dropped due to low item-total correlation |
| 10 | 10 | 55933A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 10 | 10 | 55933B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 10 | 11 | 48722 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 48739 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 11 | 48745 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses |
| 10 | 11 | 49180 | CAT | 1 | Dropped as suggested by content review |
| 10 | 11 | 49398 | CAT | 2 | Dropped as suggested by content review |
| 10 | 11 | 49452 | CAT | 2 | Dropped as suggested by content review |
| 10 | 11 | 49460 | CAT | 2 | Dropped as suggested by content review |
| 10 | 11 | 49585 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses |
| 10 | 11 | 49631 | CAT | 3 | Dropped due to no scored responses |
| 10 | 11 | 49635 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 11 | 49675 | CAT | 3 | Dropped as suggested by content review |
| 10 | 11 | 55150 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55151 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55153A | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55153B | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55153C | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55154 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55155 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55156 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55157A | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55157B | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55157C | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55158 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55159 | PT | 1 | Dropped due to no scored responses |
| 10 | 11 | 55160 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55161 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55162A | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55162B | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55162C | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55164 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55165 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55166A | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55166B | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55166C | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55604 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55921 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses and sparse responses |
| 10 | 11 | 55928 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55929 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55934 | PT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 10 | 11 | 55935 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55936A | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55936B | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55936C | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55937 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55938 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55939 | PT | 4 | Dropped due to no scored responses |
| 10 | 11 | 55940A | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55940B | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 55940C | PT | 2 | Dropped due to no scored responses |
| 10 | 11 | 56097 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 46724 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 46726 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 46728 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 47779 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 47787 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 47789 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 48055 | CAT | 2 | Dropped as suggested by content review |
| 11 | 9 | 48067 | CAT | 2 | Dropped as suggested by content review |
| 11 | 9 | 48259 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 48607 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 53033 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53034 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53035 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53036 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53037A | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53037B | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53037C | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53058 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53059 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53060 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 53061A | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53061B | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53061C | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 53378 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 9 | 53390 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 53392 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 53416 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 9 | 53419 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 9 | 53421 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 53435 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 53439 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 53473 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 53488 | CAT | 3 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 9 | 53490 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 53492 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 53630 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 9 | 55091 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55096 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55102 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55108A | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55108B | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55108C | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55111 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55112 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55113 | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55114A | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55114B | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55114C | PT | 1 | Dropped due to no scored responses |
| 11 | 9 | 55553 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55556 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55557 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55559A | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55559B | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55559C | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55598 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55600 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55601 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55624 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55625 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55626 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55627A | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55627B | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55627C | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55902 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55903 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55904 | PT | 4 | Dropped due to no scored responses |
| 11 | 9 | 55905A | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55905B | PT | 2 | Dropped due to no scored responses |
| 11 | 9 | 55905C | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 48704 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 48719 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 48726 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 48846 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 48897 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 48907 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 48909 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 48923 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 10 | 48925 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 49002 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses and sparse responses |
| 11 | 10 | 49356 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 49530 | CAT | 4 | Dropped due to low item-total correlation |
| 11 | 10 | 49532 | CAT | 4 | Dropped due to low item-total correlation |
| 11 | 10 | 49536 | CAT | 4 | Dropped as suggested by content review |
| 11 | 10 | 49599 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 49603 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 51542 | CAT | 2 | Dropped due to no scored responses |
| 11 | 10 | 51554 | CAT | 2 | Dropped due to no scored responses |
| 11 | 10 | 53517 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 53530 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 53538 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 53548 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses and sparse responses |
| 11 | 10 | 53558 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 53594 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 53596 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 53598 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 53600 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 53606 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 53612 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 53620 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 10 | 53624 | CAT | 4 | Dropped as suggested by content review |
| 11 | 10 | 53628 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses |
| 11 | 10 | 55097 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55101 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55104 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55107A | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55107B | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55107C | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55258 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55259 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55260A | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55260B | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55260C | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55619 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55620 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55621 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55622 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55623A | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55623B | PT | 2 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 10 | 55623C | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55918 | PT | 4 | Dropped due to low item-total correlation |
| 11 | 10 | 55930 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55931 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55932 | PT | 4 | Dropped due to no scored responses |
| 11 | 10 | 55933A | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55933B | PT | 2 | Dropped due to no scored responses |
| 11 | 10 | 55933C | PT | 2 | Dropped due to no scored responses |
| 11 | 11 | 48705 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 48709 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 11 | 11 | 48712 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 48722 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 48739 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 48745 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses |
| 11 | 11 | 49133 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49139 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49180 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 49190 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49198 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses and sparse responses |
| 11 | 11 | 49200 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to nonmonotonic responses |
| 11 | 11 | 49228 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49236 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49238 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49392 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 49398 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49408 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49420 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49448 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49452 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49460 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49468 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49472 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 49502 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49559 | CAT | 4 | Dropped as suggested by content review |
| 11 | 11 | 49635 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 49657 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 49675 | CAT | 3 | Dropped as suggested by content review |
| 11 | 11 | 50204 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 54367 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 55153A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| | | | | | responses |
| 11 | 11 | 55153B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55157A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55157B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55162A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55162B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55166A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 11 | 11 | 55166B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 11 | 11 | 55921 | PT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 55936A | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55936B | PT | 4 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,3 due to sparse responses |
| 11 | 11 | 55940A | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |
| 11 | 11 | 55940B | PT | 2 | Collapsed categories: 0,1,2,3,4 becomes 0,0,1,2,2 due to sparse responses |

Table B.2. Math Items Receiving Pre-treatment before Calibration based on Data Clearning Procedure

| Admin Grade | Item Grade | Item Number | CAT/PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 3 | 3 | 47167 | CAT | 3 | Dropped as suggested by content review |
| 3 | 3 | 47185 | CAT | 3 | Collapsed categories: 0,1,2,3 becomes 0,1,1,2 due to nonmonotonic responses |
| 3 | 3 | 48754 | CAT | 3 | Dropped as suggested by content review |
| 3 | 3 | 51720 | CAT | 1 | Dropped as suggested by content review |
| 3 | 3 | 51728 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 51756 | CAT | 1 | Dropped due to low item-total correlation |
| 3 | 3 | 51806 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 3 | 4 | 45981 | CAT | 1 | Dropped as suggested by content review |
| 3 | 4 | 45983 | CAT | 1 | Dropped as suggested by content review |
| 3 | 4 | 51666 | CAT | 1 | Dropped as suggested by content review |
| 3 | 4 | 51672 | CAT | 1 | Dropped as suggested by content review |
| 3 | 4 | 51684 | CAT | 1 | Dropped due to no scored responses |
| 3 | 4 | 51686 | CAT | 1 | Dropped due to no scored responses |
| 3 | 4 | 51694 | CAT | 2 | Dropped due to no scored responses |
| 3 | 4 | 51696 | CAT | 1 | Dropped due to no scored responses |
| 3 | 4 | 51700 | CAT | 1 | Dropped due to no scored responses |
| 3 | 4 | 51702 | CAT | 2 | Dropped due to no scored responses |
| 4 | 3 | 47167 | CAT | 3 | Dropped as suggested by content review |
| 4 | 3 | 48754 | CAT | 3 | Dropped as suggested by content review |
| 4 | 3 | 51756 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 3 | 51816 | CAT | 1 | Dropped due to no scored responses |
| 4 | 3 | 51822 | CAT | 1 | Dropped due to no scored responses |
| 4 | 3 | 51832 | CAT | 2 | Dropped due to no scored responses |
| 4 | 3 | 51834 | CAT | 2 | Dropped due to no scored responses |
| 4 | 3 | 51856 | CAT | 3 | Dropped due to no scored responses |
| 4 | 4 | 45981 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 45983 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 51666 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 51672 | CAT | 1 | Dropped as suggested by content review |
| 4 | 4 | 53177 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 5 | 53303 | CAT | 1 | Dropped due to low item-total correlation |
| 4 | 5 | 53760 | CAT | 1 | Dropped due to no scored responses |
| 4 | 5 | 54341 | CAT | 1 | Dropped due to no scored responses |
| 4 | 5 | 54363 | CAT | 1 | Dropped due to no scored responses |
| 4 | 5 | 54955 | CAT | 4 | Dropped due to no scored responses |
| 5 | 4 | 53147 | CAT | 1 | Dropped due to no scored responses |
| 5 | 4 | 53151 | CAT | 1 | Dropped due to no scored responses |
| 5 | 4 | 53155 | CAT | 1 | Dropped due to no scored responses |
| 5 | 4 | 53157 | CAT | 1 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 5 | 4 | 53159 | CAT | 1 | Dropped due to no scored responses |
| 5 | 4 | 53229 | CAT | 2 | Dropped due to no scored responses |
| 5 | 4 | 53271 | CAT | 2 | Dropped due to no scored responses |
| 5 | 5 | 42994 | CAT | 3 | Dropped as suggested by content review |
| 5 | 5 | 43574 | CAT | 1 | Dropped as suggested by content review |
| 5 | 5 | 45625 | CAT | 4 | Dropped due to low item-total correlation |
| 5 | 5 | 45967 | CAT | 2 | Dropped due to low item-total correlation |
| 5 | 5 | 47158 | CAT | 1 | Dropped as suggested by content review |
| 5 | 5 | 53101 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 5 | 5 | 53303 | CAT | 1 | Dropped due to low item-total correlation |
| 5 | 5 | 54365 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 5 | 5 | 54955 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 5 | 6 | 42925 | CAT | 1 | Dropped as suggested by content review |
| 5 | 6 | 43482 | CAT | 2 | Dropped as suggested by content review |
| 5 | 6 | 46647 | CAT | 3 | Dropped due to no scored responses |
| 5 | 6 | 46938 | CAT | 1 | Dropped due to low item-total correlation |
| 5 | 6 | 48655 | CAT | 2 | Dropped as suggested by content review |
| 5 | 6 | 52924 | CAT | 1 | Dropped due to no scored responses |
| 5 | 6 | 52930 | CAT | 1 | Dropped due to no scored responses |
| 6 | 5 | 42994 | CAT | 3 | Dropped as suggested by content review |
| 6 | 5 | 43574 | CAT | 1 | Dropped as suggested by content review |
| 6 | 5 | 45625 | CAT | 4 | Dropped due to low item-total correlation |
| 6 | 5 | 45967 | CAT | 2 | Dropped due to low item-total correlation |
| 6 | 5 | 52617 | CAT | 1 | Dropped due to no scored responses |
| 6 | 5 | 53095 | CAT | 1 | Dropped due to no scored responses |
| 6 | 5 | 53097 | CAT | 1 | Dropped due to no scored responses |
| 6 | 5 | 53099 | CAT | 1 | Dropped due to no scored responses |
| 6 | 5 | 53103 | CAT | 2 | Dropped due to no scored responses |
| 6 | 6 | 42781 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 42925 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 42986 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 43209 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 43383 | CAT | 2 | Dropped as suggested by content review |
| 6 | 6 | 43482 | CAT | 2 | Dropped as suggested by content review |
| 6 | 6 | 43787 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 43910 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 44038 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 44055 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 46045 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 46049 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 46558 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 46652 | CAT | 4 | Dropped as suggested by content review |
| 6 | 6 | 46938 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 46944 | CAT | 1 | Dropped due to low item-total correlation |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 6 | 6 | 46962 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 6 | 47032 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 48280 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 48575 | CAT | 2 | Dropped as suggested by content review |
| 6 | 6 | 48581 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 48649 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 6 | 6 | 48655 | CAT | 2 | Dropped as suggested by content review |
| 6 | 6 | 48663 | CAT | 1 | Dropped as suggested by content review |
| 6 | 6 | 51411 | CAT | 3 | Dropped as suggested by content review |
| 6 | 6 | 52945 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,0,1 due to nonmonotonic responses |
| 6 | 7 | 42983 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 6 | 7 | 44766 | CAT | 2 | Dropped as suggested by content review |
| 6 | 7 | 44824 | CAT | 1 | Dropped as suggested by content review |
| 6 | 7 | 45059 | CAT | 1 | Dropped as suggested by content review |
| 6 | 7 | 45726 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 7 | 46069 | CAT | 1 | Dropped due to no scored responses |
| 6 | 7 | 46761 | CAT | 1 | Dropped due to no scored responses |
| 6 | 7 | 46834 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 7 | 46844 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 7 | 46854 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 7 | 46868 | CAT | 1 | Dropped due to low item-total correlation |
| 6 | 7 | 48075 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 42781 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 42872 | CAT | 1 | Dropped due to no scored responses |
| 7 | 6 | 42873 | CAT | 1 | Dropped due to no scored responses |
| 7 | 6 | 43383 | CAT | 2 | Dropped as suggested by content review |
| 7 | 6 | 46962 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 6 | 47032 | CAT | 1 | Dropped as suggested by content review |
| 7 | 6 | 48575 | CAT | 2 | Dropped as suggested by content review |
| 7 | 6 | 48577 | CAT | 1 | Dropped due to no scored responses |
| 7 | 6 | 48579 | CAT | 1 | Dropped due to no scored responses |
| 7 | 6 | 48581 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 42746 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 43394 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 43515 | CAT | 4 | Dropped due to low item-total correlation |
| 7 | 7 | 43522 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 43524 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 44271 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 44766 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 44785 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 44823 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 44824 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 44828 | CAT | 4 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 7 | 7 | 44831 | CAT | 3 | Dropped as suggested by content review |
| 7 | 7 | 44832 | CAT | 3 | Dropped as suggested by content review |
| 7 | 7 | 44906 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45059 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45090 | CAT | 4 | Dropped as suggested by content review |
| 7 | 7 | 45093 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45103 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45105 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45581 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 45726 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46061 | CAT | 2 | Dropped due to low item-total correlation |
| 7 | 7 | 46785 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 7 | 7 | 46816 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46818 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46824 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46826 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46834 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46838 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46840 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46842 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46844 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46856 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46864 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46866 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 46868 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46872 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46886 | CAT | 1 | Dropped as suggested by content review |
| 7 | 7 | 46902 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 7 | 46918 | CAT | 2 | Dropped as suggested by content review |
| 7 | 7 | 48075 | CAT | 1 | Dropped as suggested by content review |
| 7 | 8 | 42807 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 7 | 8 | 42942 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 7 | 8 | 43529 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 7 | 8 | 44193 | CAT | 2 | Dropped as suggested by content review |
| 7 | 8 | 44834 | CAT | 3 | Dropped as suggested by content review |
| 7 | 8 | 45993 | CAT | 1 | Dropped due to no scored responses |
| 7 | 8 | 46803 | CAT | 1 | Dropped as suggested by content review |
| 7 | 8 | 46807 | CAT | 1 | Dropped due to no scored responses |
| 7 | 8 | 47010 | CAT | 1 | Dropped due to low item-total correlation |
| 7 | 8 | 47092 | CAT | 1 | Dropped as suggested by content review |
| 7 | 8 | 48113 | CAT | 1 | Dropped due to no scored responses |
| 8 | 7 | 42867 | CAT | 1 | Dropped due to no scored responses |
| 8 | 7 | 43394 | CAT | 2 | Dropped as suggested by content review |
| 8 | 7 | 43524 | CAT | 1 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 8 | 7 | 44271 | CAT | 1 | Dropped as suggested by content review |
| 8 | 7 | 44823 | CAT | 1 | Dropped as suggested by content review |
| 8 | 7 | 44828 | CAT | 4 | Dropped as suggested by content review |
| 8 | 7 | 44831 | CAT | 3 | Dropped as suggested by content review |
| 8 | 7 | 44832 | CAT | 3 | Dropped as suggested by content review |
| 8 | 7 | 45090 | CAT | 4 | Dropped as suggested by content review |
| 8 | 7 | 45581 | CAT | 1 | Dropped as suggested by content review |
| 8 | 7 | 46061 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 7 | 46783 | CAT | 1 | Dropped due to no scored responses |
| 8 | 7 | 46818 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 7 | 46886 | CAT | 1 | Dropped as suggested by content review |
| 8 | 7 | 46918 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 43118 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 43389 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 43814 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 43815 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 44051 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 44192 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 44193 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 44194 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 44660 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 44664 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 44789 | CAT | 4 | Dropped as suggested by content review |
| 8 | 8 | 44834 | CAT | 3 | Dropped as suggested by content review |
| 8 | 8 | 44865 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 45001 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 45002 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45004 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 45008 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45020 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 45021 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 45717 | CAT | 2 | Dropped as suggested by content review |
| 8 | 8 | 45859 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45871 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45877 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45881 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 45887 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 46013 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 8 | 46055 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 46057 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 8 | 46077 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 8 | 46570 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 8 | 8 | 46803 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47010 | CAT | 1 | Dropped due to low item-total correlation |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 8 | 8 | 47092 | CAT | 1 | Dropped as suggested by content review |
| 8 | 8 | 47098 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 42868 | CAT | 1 | Dropped due to no scored responses |
| 8 | 9 | 43135 | CAT | 2 | Dropped due to no scored responses |
| 8 | 9 | 43683 | CAT | 4 | Dropped as suggested by content review |
| 8 | 9 | 43704 | CAT | 2 | Dropped as suggested by content review |
| 8 | 9 | 43741 | CAT | 2 | Dropped as suggested by content review |
| 8 | 9 | 43998 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 8 | 9 | 44237 | CAT | 2 | Dropped due to low item-total correlation |
| 8 | 9 | 44373 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 44383 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 44476 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 44484 | CAT | 1 | Dropped as suggested by content review |
| 8 | 9 | 44938 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 44975 | CAT | 3 | Dropped as suggested by content review |
| 8 | 9 | 45095 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 45404 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 45428 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 45438 | CAT | 1 | Dropped as suggested by content review |
| 8 | 9 | 45478 | CAT | 1 | Dropped as suggested by content review |
| 8 | 9 | 45485 | CAT | 1 | Dropped as suggested by content review |
| 8 | 9 | 45491 | CAT | 1 | Dropped due to no scored responses |
| 8 | 9 | 45497 | CAT | 1 | Dropped due to low item-total correlation |
| 8 | 9 | 45558 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 42864 | CAT | 1 | Dropped due to no scored responses |
| 9 | 8 | 42865 | CAT | 1 | Dropped due to no scored responses |
| 9 | 8 | 43118 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 43389 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 43520 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 8 | 44194 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 8 | 44865 | CAT | 1 | Dropped as suggested by content review |
| 9 | 8 | 44872 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 9 | 8 | 45021 | CAT | 2 | Dropped as suggested by content review |
| 9 | 8 | 45041 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 8 | 45859 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 8 | 46013 | CAT | 2 | Dropped due to low item-total correlation |
| 9 | 8 | 46033 | CAT | 1 | Dropped due to no scored responses |
| 9 | 8 | 46057 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 8 | 46077 | CAT | 2 | Dropped due to low item-total correlation |
| 9 | 8 | 46570 | CAT | 2 | Dropped due to no scored responses |
| 9 | 8 | 48119 | CAT | 2 | Dropped due to no scored responses |
| 9 | 9 | 43683 | CAT | 4 | Dropped as suggested by content review |
| 9 | 9 | 43704 | CAT | 2 | Dropped as suggested by content review |
| 9 | 9 | 43741 | CAT | 2 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 9 | 9 | 44237 | CAT | 2 | Dropped due to low item-total correlation |
| 9 | 9 | 44373 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44383 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44388 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 44444 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44475 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44484 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 44489 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 9 | 9 | 44938 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44952 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 44959 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 44972 | CAT | 3 | Dropped as suggested by content review |
| 9 | 9 | 44975 | CAT | 3 | Dropped as suggested by content review |
| 9 | 9 | 45095 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45098 | CAT | 3 | Dropped due to low item-total correlation |
| 9 | 9 | 45396 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 45400 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45402 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45404 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45410 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45426 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45428 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45438 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 45450 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45478 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 45480 | CAT | 2 | Dropped as suggested by content review |
| 9 | 9 | 45485 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 45497 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 9 | 45558 | CAT | 1 | Dropped as suggested by content review |
| 9 | 9 | 45578 | CAT | 2 | Dropped due to low item-total correlation |
| 9 | 10 | 42950 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 10 | 43101 | CAT | 1 | Dropped due to no scored responses |
| 9 | 10 | 43535 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 43543 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 43687 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 43701 | CAT | 2 | Dropped as suggested by content review |
| 9 | 10 | 43859 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 9 | 10 | 44517 | CAT | 1 | Dropped as suggested by content review |
| 9 | 10 | 44525 | CAT | 1 | Dropped due to low item-total correlation |
| 9 | 10 | 44737 | CAT | 2 | Collapsed categories: 0,1,2,3 becomes 0,1,1,1 due to sparse responses |
| 9 | 10 | 45643 | CAT | 1 | Dropped due to no scored responses |
| 9 | 10 | 45917 | CAT | 1 | Dropped due to no scored responses |
| 9 | 10 | 46611 | CAT | 1 | Dropped due to no scored responses |
| 9 | 10 | 46613 | CAT | 1 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 9 | 10 | 47026 | CAT | 1 | Dropped as suggested by content review |
| 9 | 10 | 48788 | CAT | 1 | Dropped due to no scored responses |
| 9 | 10 | 48794 | CAT | 1 | Dropped due to no scored responses |
| 10 | 9 | 44128 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 44388 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 44444 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 44475 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 44478 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 44489 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 10 | 9 | 44944 | CAT | 2 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 10 | 9 | 44948 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 44952 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 44958 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 9 | 44959 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 44972 | CAT | 3 | Dropped as suggested by content review |
| 10 | 9 | 45098 | CAT | 3 | Dropped due to low item-total correlation |
| 10 | 9 | 45396 | CAT | 1 | Dropped as suggested by content review |
| 10 | 9 | 45400 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 45402 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 45410 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 45426 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 45450 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 9 | 45480 | CAT | 2 | Dropped as suggested by content review |
| 10 | 9 | 45489 | CAT | 1 | Dropped due to no scored responses |
| 10 | 9 | 45578 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 10 | 43535 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 43543 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 43687 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 43701 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 44517 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 44525 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 44735 | CAT | 2 | Dropped as suggested by content review |
| 10 | 10 | 44737 | CAT | 2 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 10 | 10 | 45077 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 45079 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 45645 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 10 | 46988 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 46994 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 47026 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 47080 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 49776 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 10 | 49778 | CAT | 1 | Dropped as suggested by content review |
| 10 | 10 | 49780 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 11 | 42724 | CAT | 1 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 10 | 11 | 42734 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 42770 | CAT | 1 | Dropped as suggested by content review |
| 10 | 11 | 43795 | CAT | 1 | Dropped as suggested by content review |
| 10 | 11 | 43799 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 11 | 44124 | CAT | 2 | Dropped as suggested by content review |
| 10 | 11 | 44172 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 44687 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 11 | 44787 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 11 | 46089 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 10 | 11 | 49782 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49784 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49796 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49814 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 49836 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49840 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49844 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49848 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 49850 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 50049 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 11 | 50053 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 50055 | CAT | 4 | Dropped due to no scored responses |
| 10 | 11 | 50071 | CAT | 4 | Dropped due to no scored responses |
| 10 | 11 | 50079 | CAT | 2 | Dropped due to low item-total correlation |
| 10 | 11 | 50081 | CAT | 1 | Dropped due to no scored responses |
| 10 | 11 | 50089 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 50103 | CAT | 4 | Dropped due to no scored responses |
| 10 | 11 | 50125 | CAT | 1 | Dropped due to low item-total correlation |
| 10 | 11 | 50129 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 42868 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 43135 | CAT | 2 | Dropped due to no scored responses |
| 11 | 9 | 43683 | CAT | 4 | Dropped as suggested by content review |
| 11 | 9 | 43704 | CAT | 2 | Dropped as suggested by content review |
| 11 | 9 | 43741 | CAT | 2 | Dropped as suggested by content review |
| 11 | 9 | 43998 | CAT | 3 | Dropped due to no scored responses |
| 11 | 9 | 44001 | CAT | 3 | Dropped due to no scored responses |
| 11 | 9 | 44128 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 9 | 44237 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 9 | 44367 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44373 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44382 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44383 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44388 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 44444 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44473 | CAT | 4 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 9 | 44474 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44475 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44476 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44484 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 44488 | CAT | 2 | Dropped due to no scored responses |
| 11 | 9 | 44489 | CAT | 1 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 9 | 44526 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44742 | CAT | 3 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 9 | 44938 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44942 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44944 | CAT | 2 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 9 | 44950 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44952 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 44953 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44957 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44958 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 9 | 44959 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 44961 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 44972 | CAT | 3 | Dropped as suggested by content review |
| 11 | 9 | 44975 | CAT | 3 | Dropped as suggested by content review |
| 11 | 9 | 45029 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45030 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45095 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45098 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 9 | 45390 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45396 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 45400 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45402 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45404 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45410 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45426 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45428 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45438 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 45448 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45450 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45477 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45478 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 45480 | CAT | 2 | Dropped as suggested by content review |
| 11 | 9 | 45485 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 45489 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45491 | CAT | 1 | Dropped due to no scored responses |
| 11 | 9 | 45497 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 9 | 45558 | CAT | 1 | Dropped as suggested by content review |
| 11 | 9 | 45578 | CAT | 2 | Dropped due to low item-total correlation |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 10 | 42950 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 43101 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 43112 | CAT | 2 | Dropped due to no scored responses |
| 11 | 10 | 43115 | CAT | 4 | Dropped due to no scored responses |
| 11 | 10 | 43414 | CAT | 3 | Dropped due to no scored responses |
| 11 | 10 | 43489 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 43535 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 43538 | CAT | 3 | Dropped due to no scored responses |
| 11 | 10 | 43543 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 43687 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 43701 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 43740 | CAT | 2 | Dropped due to no scored responses |
| 11 | 10 | 43859 | CAT | 4 | Dropped due to no scored responses |
| 11 | 10 | 44349 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 44443 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 44517 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 44525 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 44604 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 44610 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 44615 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 44735 | CAT | 2 | Dropped as suggested by content review |
| 11 | 10 | 44737 | CAT | 2 | Collapsed categories: 0,1,2,3 becomes 0,1,1,1 due to sparse responses |
| 11 | 10 | 44774 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 44776 | CAT | 3 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 10 | 45066 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45067 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45068 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45069 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45071 | CAT | 4 | Dropped due to no scored responses |
| 11 | 10 | 45074 | CAT | 3 | Dropped due to no scored responses |
| 11 | 10 | 45075 | CAT | 4 | Collapsed categories: 0,1,2,3 becomes 0,1,1,1 due to sparse responses |
| 11 | 10 | 45077 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 45079 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 45083 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 10 | 45084 | CAT | 3 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 10 | 45089 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45463 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45466 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45470 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45643 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 45645 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 10 | 45917 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 46611 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 46613 | CAT | 1 | Dropped due to no scored responses |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 10 | 46924 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 46950 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 46952 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 46988 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 46994 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 47018 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 47026 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 47046 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 47080 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 48788 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 48794 | CAT | 1 | Dropped due to no scored responses |
| 11 | 10 | 49776 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 10 | 49778 | CAT | 1 | Dropped as suggested by content review |
| 11 | 10 | 49780 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 42770 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43170 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 43287 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43308 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43641 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 43681 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43762 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43795 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43877 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43879 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 43950 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 43954 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 44113 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 44117 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 44124 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 44155 | CAT | 2 | Dropped as suggested by content review |
| 11 | 11 | 44172 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 44322 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 44557 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 44639 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 44687 | CAT | 2 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 44693 | CAT | 4 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 44841 | CAT | 1 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 45846 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 45849 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 46233 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 47040 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 47086 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 47110 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 47188 | CAT | 3 | Dropped as suggested by content review |

| Admin Grade | Item Grade | Item Number | CAT/PT | Claim | Pre-Treatment |
|---|---|---|---|---|---|
| 11 | 11 | 48135 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 48585 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 48591 | CAT | 3 | Collapsed categories: 0,1,2 becomes 0,1,1 due to sparse responses |
| 11 | 11 | 48593 | CAT | 3 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 11 | 48665 | CAT | 3 | Dropped due to low item-total correlation |
| 11 | 11 | 48668 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 48676 | CAT | 4 | Collapsed categories: 0,1,2,3 becomes 0,1,2,2 due to sparse responses |
| 11 | 11 | 49808 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49814 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49832 | CAT | 1 | Dropped as suggested by content review |
| 11 | 11 | 49842 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49850 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49862 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49866 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 49872 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 50043 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 50049 | CAT | 2 | Dropped due to low item-total correlation |
| 11 | 11 | 50089 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 50125 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 50129 | CAT | 1 | Dropped due to low item-total correlation |
| 11 | 11 | 50133 | CAT | 1 | Dropped due to low item-total correlation |

Table B.3. ELA Items Receiving Treatment during IRT Calibration Under Different Model Combinations

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Treatment | 1PL/PC | 2PL/GPC | 3PL/GPC |
|---|---|---|---|---|---|---|---|---|
| 3 | 3 | 54269 | CAT | 4 | Category starting value re-assigned | | X | X |
| 3 | 3 | 53995 | CAT | 3 | Guessing fixed at 0.25 | | | X |
| 3 | 3 | 53977 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 3 | 3 | 54045 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 3 | 3 | 54197 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 3 | 3 | 54209 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 3 | 3 | 54287 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 3 | 4 | 54382 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 3 | 4 | 54998 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 3 | 4 | 55002 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 3 | 4 | 55011 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 4 | 3 | 54097 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 56188B | PT | 2 | Dropped due to LID | | X | X |
| 4 | 4 | 54426 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 4 | 4 | 54998 | CAT | 3 | Guessing fixed at 0.25 | | | X |
| 4 | 4 | 55011 | CAT | 3 | Guessing fixed at 0.25 | | | X |
| 4 | 4 | 55002 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55008 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55392 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 54430 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 54492 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55277 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55362 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55316 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 55426 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 5 | 54726 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 4 | 54498 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 5 | 4 | 55316 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 55110B | PT | 2 | Dropped due to LID | | X | X |
| 5 | 5 | 55547A | PT | 2 | Dropped due to LID | | X | X |
| 5 | 5 | 56274B | PT | 2 | Dropped due to LID | | X | X |
| 5 | 5 | 56322A | PT | 2 | Dropped due to LID | | X | X |
| 5 | 5 | 54726 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 5 | 5 | 54678 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54884 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54698 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54714 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54776 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54808 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54818 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 54918 | CAT | 3 | Guessing starting at 0.25 | | | X |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Treatment | 1PL/PC | 2PL/GPC | 3PL/GPC |
|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 54924 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53646 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 6 | 52843 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 5 | 52321 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 5 | 53646 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52398B | PT | 2 | Dropped due to LID | | X | X |
| 6 | 6 | 53024A | PT | 2 | Dropped due to LID | | X | X |
| 6 | 6 | 55088A | PT | 2 | Dropped due to LID | | X | X |
| 6 | 6 | 55103A | PT | 2 | Dropped due to LID | | X | X |
| 6 | 6 | 55927B | PT | 4 | Dropped due to LID | | X | X |
| 6 | 6 | 52847 | CAT | 3 | Guessing fixed at 0.25 | | | X |
| 6 | 6 | 52837 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52839 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52845 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52349 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52633 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52639 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52675 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52677 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52679 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52750 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52768 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52776 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52863 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52871 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 52877 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 47830 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 6 | 53024B | PT | 2 | Category starting value re-assigned | | | X |
| 6 | 7 | 47525 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 7 | 47557 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 7 | 6 | 47830 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 7 | 7 | 52587A | PT | 2 | Dropped due to LID | | X | X |
| 7 | 7 | 53019B | PT | 2 | Dropped due to LID | | X | X |
| 7 | 7 | 53028A | PT | 2 | Dropped due to LID | | X | X |
| 7 | 7 | 53032B | PT | 2 | Dropped due to LID | | X | X |
| 7 | 7 | 53129B | PT | 2 | Dropped due to LID | | X | X |
| 7 | 7 | 46480 | CAT | 3 | Category starting value re-assigned | | X | X |
| 7 | 7 | 55632 | PT | 4 | Category starting value re-assigned | | X | X |
| 7 | 7 | 47467 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 7 | 7 | 47557 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 7 | 7 | 47517 | CAT | 1 | Guessing starting at 0.25 | | | X |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Treatment | 1PL/PC | 2PL/GPC | 3PL/GPC |
|---|---|---|---|---|---|---|---|---|
| 7 | 7 | 47886 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 7 | 8 | 47795 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 8 | 7 | 55632 | PT | 4 | Category starting value re-assigned | | X | X |
| 8 | 7 | 47467 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 53041B | PT | 2 | Dropped due to LID | | X | X |
| 8 | 8 | 53045A | PT | 2 | Dropped due to LID | | X | X |
| 8 | 8 | 53050B | PT | 2 | Dropped due to LID | | X | X |
| 8 | 8 | 47319 | CAT | 3 | Guessing fixed at 0.25 | | | X |
| 8 | 8 | 53049 | PT | 4 | Guessing fixed at 0 | | | X |
| 8 | 8 | 47687 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47960 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47397 | CAT | 4 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47643 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47653 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47663 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47699 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47956 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47976 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47327 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47329 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47333 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 47968 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 8 | 8 | 53134 | PT | 4 | Guessing starting at 0.25 | | | X |
| 9 | 8 | 47329 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 9 | 8 | 47333 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 9 | 8 | 47968 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 9 | 8 | 53134 | PT | 4 | Guessing starting at 0.25 | | | X |
| 9 | 8 | 53049 | PT | 4 | Guessing fixed at 0 | | | X |
| 9 | 9 | 53037B | PT | 2 | Dropped due to LID | | X | X |
| 9 | 9 | 53061B | PT | 2 | Dropped due to LID | | X | X |
| 9 | 9 | 55108B | PT | 1 | Dropped due to LID | | X | X |
| 9 | 9 | 55559B | PT | 2 | Dropped due to LID | | X | X |
| 9 | 9 | 55627B | PT | 2 | Dropped due to LID | | X | X |
| 9 | 9 | 55114B | PT | 1 | Dropped due to LID | | X | X |
| 9 | 9 | 55905B | PT | 2 | Dropped due to LID | | X | X |
| 9 | 9 | 47785 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 9 | 9 | 53423 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 9 | 9 | 53433 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 9 | 9 | 53450 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 9 | 9 | 53372 | CAT | 1 | Guessing fixed at 0.10 | | | X |
| 9 | 9 | 46672 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 9 | 10 | 53550 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 9 | 10 | 53616 | CAT | 3 | Guessing starting at 0.25 | | | X |

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Treatment | 1PL/PC | 2PL/GPC | 3PL/GPC |
|---|---|---|---|---|---|---|---|---|
| 10 | 9 | 53372 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 10 | 9 | 53423 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 10 | 9 | 53433 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 10 | 10 | 55107B | PT | 2 | Dropped due to LID | | X | X |
| 10 | 10 | 55260B | PT | 2 | Dropped due to LID | | X | X |
| 10 | 10 | 55623B | PT | 2 | Dropped due to LID | | X | X |
| 10 | 10 | 55933B | PT | 2 | Dropped due to LID | | X | X |
| 10 | 10 | 53580 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 10 | 10 | 53614 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 10 | 10 | 53616 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 10 | 10 | 49332 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 10 | 10 | 49619 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 10 | 11 | 49280 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 10 | 11 | 49633 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 47777 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 48167 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 53372 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 53431 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 53433 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 53455 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 11 | 9 | 53486 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 11 | 10 | 53580 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 11 | 10 | 53608 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 11 | 10 | 49332 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 11 | 11 | 55153B | PT | 2 | Dropped due to LID | | X | X |
| 11 | 11 | 55940A | PT | 2 | Dropped due to LID | | X | X |
| 11 | 11 | 49384 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 11 | 11 | 49368 | CAT | 2 | Guessing starting at 0.25 | | | X |
| 11 | 11 | 49544 | CAT | 4 | Guessing starting at 0.25 | | | X |
| 11 | 11 | 49679 | CAT | 3 | Guessing starting at 0.25 | | | X |
| 11 | 11 | 49723 | CAT | 3 | Guessing starting at 0.25 | | | X |

Table B.4. Math Items Receiving Treatment during IRT Calibration Under Different Model Combinations

| Admin Grade | Item Grade | Item Number | CAT/ PT | Claim | Treatment | 1PL/PC | 2PL/GPC | 3PL/GPC |
|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 43075 | CAT | 3 | Category starting value re-assigned | | | X |
| 3 | 3 | 51730 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 3 | 3 | 51754 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 5 | 44361 | CAT | 1 | Category starting value re-assigned | | X | |
| 4 | 4 | 43097 | CAT | 1 | Category starting value re-assigned | | | X |
| 4 | 4 | 45957 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 51682 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 4 | 4 | 45959 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 45909 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 45931 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53319 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53362 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53091 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53093 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 5 | 5 | 53307 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 6 | 7 | 44822 | CAT | 1 | Category 2 and 3 merged | | X | X |
| 6 | 7 | 45052 | CAT | 3 | Category starting value re-assigned | | X | X |
| 6 | 6 | 43705 | CAT | 2 | Category starting value re-assigned | | | X |
| 6 | 6 | 47006 | CAT | 1 | Guessing starting at 0.25 | | | X |
| 6 | 7 | 42983 | CAT | 1 | Dropped due to poor statistics ($G^2$ higher than 1,000 and r-bis lower than .01) | | | X |
| 7 | 8 | 47012 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 8 | 9 | 42890 | CAT | 1 | Category starting value re-assigned | | X | |
| 8 | 7 | 44821 | CAT | 1 | Category starting value re-assigned | | | X |
| 9 | 9 | 44951 | CAT | 2 | Category starting value re-assigned | | X | X |
| 9 | 10 | 45073 | CAT | 3 | Category 3 and 4 merged | | X | X |
| 9 | 10 | 47058 | CAT | 1 | Guessing fixed at 0.10 | | | X |
| 10 | 9 | 42767 | CAT | 1 | Category starting value re-assigned | | X | |
| 10 | 11 | 43304 | CAT | 1 | Category 2 and 3 merged | | X | |
| 10 | 11 | 44843 | CAT | 1 | Category starting value re-assigned | | | X |
| 10 | 11 | 50069 | CAT | 1 | Guessing fixed at 0.00 | | | X |
| 11 | 11 | 44853 | CAT | 1 | Category starting value re-assigned | | | X |
| 11 | 11 | 50131 | CAT | 1 | Guessing fixed at 0.25 | | | X |
| 11 | 11 | 46932 | CAT | 1 | Guessing fixed at 0.25 | | | X |

Appendix K – 2014 Field Test Report

# FIELD TEST: REPORT

**Smarter Balanced** Assessment Consortium

From March 25 to June 13, 2014, more than 4.2 million students, 16,549 schools, and thousands of teachers participated in the Smarter Balanced Field Test—the largest online assessment ever.

The Field Test helped Smarter Balanced ensure that test questions are accurate and fair for all students. The Field Test also gave students, teachers, and schools an opportunity to experience the assessment under real-world conditions and prepare for administration in spring 2015. Because questions may be revised or dropped after the Field Test, students did not receive scores.

Smarter Balanced member states learned important lessons about test administration and technology readiness that will be used to improve the assessments next year. In addition, almost all member states conducted surveys of students, teachers, and administrators to elicit feedback about the test and the testing process.

Smarter Balanced is commissioning a thorough, independent review of this feedback and will release a complete report later this summer. However, based on state feedback to date, help desk calls, and other indicators, the Consortium has already identified the following lessons:

## The technology performed well

- The test delivery platform functioned well, with limited glitches given the volume of testing: approximately 184,000 simultaneous users on peak days.
- Software and system issues were identified and resolved quickly.
- Schools tested technology and bandwidth under real-world conditions and identified practical steps to eliminate glitches, such as disabling automatic software updates that could disrupt student testing.



Washington, Montana, North Dakota, New Hampshire, Vermont, Maine, Oregon, Idaho, Michigan, Wisconsin, Michigan, Connecticut, Wyoming, South Dakota, Iowa, Nevada, Delaware, California, West Virginia, Missouri, Hawaii, North Carolina, U.S. Virgin Islands

Administered Spring 2014 Field Test

## Field Test by the Numbers

**4.2** MILLION STUDENTS

**16.5** THOUSAND SCHOOLS

**12.2** MILLION TESTS COMPLETED*

**4.5 million** tests administered with accessibility features

**1,100** Help Desk inquiries per day

> "It made me want to try, since I got to explain my reasoning and explain why I chose what I chose.
>
> - 8th grade student, Idaho

*Smarter Balanced assessments consist of two parts. Each "test" reported is one of the two parts of the English language arts (ELA)/literacy or math assessment.

SmarterBalanced.org

## Maintaining test security in the age of social media can be a challenge

- Smarter Balanced worked with states to develop a successful process for finding test questions posted online by students.
- District Test Coordinators, administrators, and teachers focused on strengthening test security, and the numbers of postings decreased dramatically.

## Ongoing communication is essential

- Smarter Balanced provided schools with communications materials to reach out to parents and respond to media inquiries about the Field Test.
- After identifying areas that caused confusion in some schools, the Test Administration Manual and test system user guides were edited for clarity.
- Smarter Balanced will continue to improve these documents as we collect feedback from schools.

> "[Students] think it's way cooler to take the test online rather than the old bubble-in multiple choice."
>
> - Superintendent and principal, California

## All students can participate in online assessments

- Students could access an unprecedented number of language supports, including interactive glossaries in 10 languages and multiple dialects, as well as full Spanish translations of the math assessment.
- Students who are deaf or hard of hearing received tests in American Sign Language, signed by recorded human interpreters.
- Refreshable Braille keyboards and real-time embossers allowed students who are blind to receive their online tests in Braille.



## Next Steps

Information from the Field Test will be used this fall to establish achievement levels for the assessment system. Achievement level setting is the process for establishing one or more cut scores on an assessment, making it possible to create categories of performance. In addition, Smarter Balanced will:

✓ Review question responses and flag questions for editing or deletion

✓ Review the directions for test items and performance tasks to ensure that they are clear for teachers and students

✓ Analyze teacher and student surveys conducted by states and publish a report later this summer

✓ Launch a Usability, Accessibility, and Accommodations Guidelines Committee to review additional universal tools, designated supports, and accommodations for inclusion in the operational assessments

SmarterBalanced.org

Appendix L –     Summative Assessment Alignment Study Brief

# Summative Assessment Alignment Study: Smarter Balanced TAC Update

Brian Gong (TAC member)
with consultation by HumRRO project team

Smarter Balanced TAC Meeting
August 27, 2014 via WebEx teleconference

# Overview

- Background of Alignment Project and Purpose of Update

- Focused TAC Member Review

- Next steps

# Project Background

- Goals & Deliverables, amended
- Contractor - HumRRO
- Approach
- Progress reports to TAC
  - July 2013, November 2013, & July 2014
- Charge to contractor from 7/17/14 meeting

# Focused TAC Member Review August 2014

- Objective
- Process
- Participants (Brian Gong, Joe Ryan)
- Outcomes
  - Found analysis approach innovative, commendable
  - Noted still a work in progress
  - TAC members focused on communication
  - Suggestions incorporated by HumRRO in revised document

# Overview of HumRRO Alignment Approach and Methodology

- Variant of Webb alignment approach
- Adapted to Smarter Balanced's circumstances
  - Focus on test specifications as well as items/forms
  - Tuned to Smarter Balanced's test specifications
    - Content specification structure
    - DOK specification structure
  - CAT design
  - Smarter Balanced schedule and data availability for operational items/"forms"

# Summary of HumRRO Study
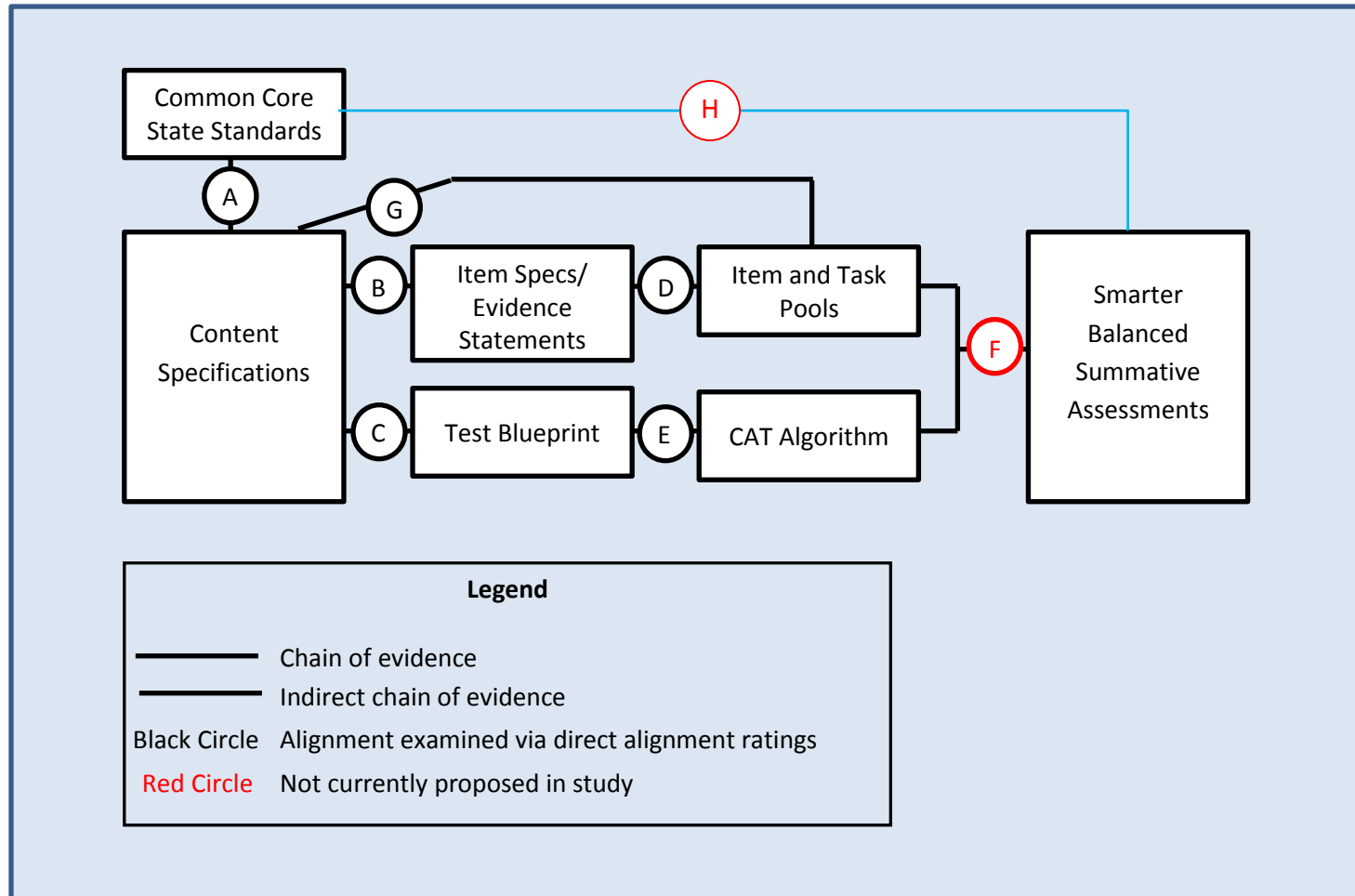
# Example: Analysis A

- Focus: Relationships between Common Core State Standards (CCSS) and Smarter Balanced Content Specifications
  - Content
  - DOK (Depth of Knowledge)

Common Core State Standards

A

Content Specifications
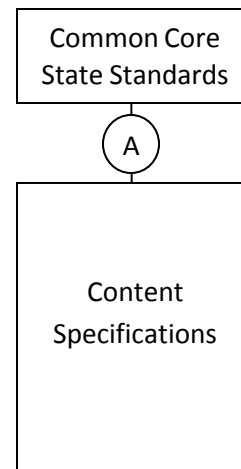
# Example: Analysis A – 2

| Common Core State Standards (Math) | Smarter Balanced Content Specs (Math) | Example from Smarter Balanced/CCSS |
|---|---|---|
| | Claim | "Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency" |
| Domain | Domain | Number and Operations/Mathematical Practices |
| Cluster | Cluster | Number and Operations—Base Ten |
| | Target | "Represent and solve problems involving multiplication and division" "Understand properties of multiplication and the relationship between multiplication and division" "Multiply and divide within 100" |
| Standard | Standard | "Interpret products of whole numbers, e.g., interpret 5 x 7 as the total number of objects in 5 groups of 7 objects each" "Understand division as an unknown-factor problem"; Fluently multiply and divide within 100, using strategies such as the relationship between multiplication and division [e.g., knowing that 8 x 5 = 40, one knows that 40 ÷ 5 = 8] or properties of operations" |

# Example: Smarter Balanced Content Specifications

**Grade**

**Claim**

**Target** – designated major [m] or supporting [a/s]

**Target text** describing assessment target

## GRADE 3 Summative Assessment Targets
## Providing Evidence Supporting Claim #1

**Claim #1: Students can explain and apply mathematical concepts and carry out mathematical procedures with precision and fluency.**

Designated **Clusters**

### Operations and Algebraic Thinking

**Target A [m]: Represent and solve problems involving multiplication and division. (DOK 1)**
Items/tasks for this target require students to use multiplication and division within 100 to solve straightforward, one-step contextual word problems in situations involving equal groups, arrays, and measurement quantities such as length, liquid volume, and masses/weights of objects. These problems should be of the equal-groups and arrays-situation types, but can include more difficult measurement quantity situations. All of these items/tasks will code straightforwardly to standard 3.OA.3. Few of these tasks coding to this standard will make the method of solution a separate target of assessment. Other tasks associated with this target will probe student understanding of the meanings of multiplication and division (3.OA.1,2). Non-contextual tasks that explicitly ask the student to determine the unknown number in a multiplication or division equation relating three whole numbers (3.OA.4) will support the development of items that provide a range of difficulty necessary for populating an adaptive item bank (see section *Understanding Assessment Targets in an Adaptive Framework,* below, for further explication).

Designated content **standards**

**Target B [m]: Understand properties of multiplication and the relationship between multiplication and division. (DOK 1)**
**Target C [m]: Multiply and divide within 100. (DOK 1)**
**Target D [m]: Solve problems involving the four operations, and identify and explain patterns in arithmetic. (DOK 2)**

### Number and Operations—Base Ten

**Target E [a/s]: Use place value understanding and properties of arithmetic to perform multi-digit arithmetic. (DOK 1)**
**Target F [m]: Develop understanding of fractions as numbers. (DOK 1, 2)**

### Measurement and Data

**Target G [m]: Solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects. (DOK 1, 2)**
**Target H [a/s]: Represent and interpret data. (DOK 2)**
**Target I [m]: Geometric measurement: understand concepts of area and relate area to multiplication and to addition. (DOK 2)**
**Target J [a/s]: Geometric measurement: recognize perimeter as an attribute of plane figures and distinguish between linear and area measures. (DOK 1)**

### Geometry

**Target K [a/s]: Reason with shapes and their attributes. (DOK 1, 2)**

Designated **DOK** – may be multiple

# Two-way alignment

- Alignment is typically analyzed in relation to a reference. Two-way alignment checks the relationship using both CCSS and the Smarter Balanced content specifications as the reference.



CCSS Content Standards

Smarter Balanced Content Specifications (Targets)

1    2    3
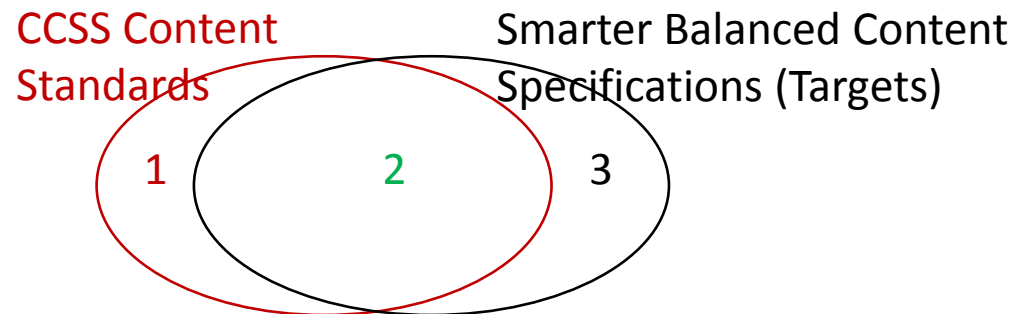
1 = CCSS content standards not aligned with Smarter Balanced Content Specifications (Targets)

2 = CCSS content standards and Smarter Balanced Content Specifications aligned with each other

3 = Smarter Balanced Content Specifications (Targets) not aligned with CCSS content standards

# Analysis A – Focus & Questions

| Focus | Questions |
|---|---|
| **Criterion: Content Representation**<br><br>The content representation (CR) criteria examine how well the content in the CCSS are represented by the assessment Targets. The CR investigations are focused on the following six questions: | Question A.CR-1. Do the grade-level standards collectively reflect the content and skills required by the target?<br><br>Question A.CR-2. Do the targets collectively reflect the content and skills required by the grade-level standard?<br><br>Question A.CR-3. Do the individual grade-level standards reflect the content and skills required by the intended targets?<br><br>Question A.CR-4: Do the individual targets reflect the content and skills required by the intended grade-level standard?<br><br>Question A.CR-5. Does each mathematical practice reflect skills required by the intended target?<br><br>Question A.CR-6. Do the reviewers agree with the intended mapping of targets and grade-level standards as identified in the content specifications? |
| **Criterion: DOK Distribution**<br><br>The DOK distribution (DD) criteria examines the reviewers' DOK distribution of the targets compared to the DOK distribution identified in the Smarter Balanced content specifications. The DD investigations are focused on the following three questions: | Question A.DD-1. Does the DOK distribution of the targets identified by the reviewers match that of the distribution identified in the content specifications (using the max DOK level)?<br><br>Question A.DD-2. Does the DOK distribution of the targets identified by the reviewers match that of the distribution identified in the content specifications (using the each independent DOK level)?<br><br>Question A.DD-3. Do the reviewers agree with the intended target DOK levels as identified in the content specifications? |

# Example Analysis A Question and Specific Methodology

Question A.CR-1. Do the grade-level standards collectively reflect the the content and skills required by the target?

**Analysis**: Compute the mean percentage of targets that were rated holistically as (a) fully-aligned (target was adequately measured across all aligned grade-level standards), (b) mostly-aligned, (c) somewhat-aligned, and (d) small portion aligned

Step 1. For each reviewer, compute the percentage of targets that were rated holistically as (a) fully-aligned, (b) mostly-aligned, (c) somewhat-aligned, and (d) small portion aligned to the full set of grade-level standards

Step 2. For each claim, compute the average percentage for each alignment rating (e.g., fully-aligned, mostly-aligned) across reviewers

Available Data:      Reviewers' holistic target coverage ratings (how well the target was represented by all of the grade-level standards identified by reviewers as being aligned to that target (4-point scale)

Table Example 1. Mean Percentage of Targets at Each Holistic Rating *(made-up data)*

| Grade | Claim | Holistic Target Rating | | | |
| | | Fully-aligned | Mostly-aligned | Somewhat-aligned | Small-portion aligned |
|---|---|---|---|---|---|
| 3 | 1 | 80% | 15% | 5% | 0% |
| | 2 | 85% | 10% | 3% | 2% |
| | 3 | 90% | 5% | 5% | 0% |
| | 4 | 90% | 5% | 5% | 0% |

# Next Steps

- Application of approach to analyze alignment of Smarter Balanced summative assessments (via test specifications) Sept./Oct. 2014

- Report submitted to Smarter Balanced by Oct. 2014

- Reviews, modifications, approval, dissemination by end of Dec. 2014

- Smarter Balanced responsible for completing additional alignment studies after operational items and "forms" available

# Comments by TAC

Appendix M –    Test Validation Worksheet

| Activity | Content | Responses | Structure | External | Conseq | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | Contract Number | Evidence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Evidence Source:** | | | | | **Purpose:** | | | | | | | **Purpose:** | | | | **Purpose:** | | | | | | |
| 1 Audit of test construction practices | √ | | | √ | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | | √ | √ | √ | 5 | Test blueprints, algorithm, item bank summaries, simulations |
| 2 Analysis of measurement precision | | | | √ | | | √ | √ | √ | √ | √ | √ | | | | | | | | | | 5 | Tech Manual, SEM FT reports, simulations |
| 3 Audit of test administration | √ | | | √ | √ | √ | | | | | | √ | | | √ | | √ | | | | | 19 | Report of administration |
| 4 Evaluation of scoring | √ | | √ | | | √ | √ | √ | √ | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | 5, 17 | AI scoring research and report; report of human scoring processes, reliability |
| 5 Analysis of scaling and equating | | | √ | | | | √ | √ | | √ | √ | | √ | | | | | | | | | 5 | Pilot analysis of dimensionality study and choice of IRT model. |
| 6 Evaluation of standard setting | √ | | √ | √ | | √ | √ | √ | √ | √ | √ | √ | | | √ | | √ | | | √ | | 21 | Report of ALD development; Standard setting plan; Final report of standard setting. |
| 7 Evaluation of fairness | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | 4, 5, 6, 8, 12, 13, 14, 16, 17, 19, 21 | Evidence from all contracts: Item development and review; DIF analysis for pilot and field test; sampling and recruitment; scoring processes and monitoring; standard setting |
| 8 Evaluation of equitable particp. & access | √ | | | | √ | √ | | | | | | √ | | | √ | | | | | √ | √ | 4, 5, 6, 8, 13, 17, 19 | Admin report of participation and use of accommodations, recuitment and sampling, scoring analysis. |
| 9 Audit of test security | | | √ | √ | | √ | √ | √ | | | √ | | √ | | | | | | | | | 19 | Test delivery and administration report |
| 10 Content validity and alignment | √ | | | | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | 14, 16 | Alignment report from contractor-item alignment to standards, test balance; analysis of blueprint fidelity |
| 11 Evaluating ECD | √ | √ | | | | √ | √ | √ | √ | | | | √ | √ | √ | | | | | | | 14, 16 | Report of item development activities; test blueprints |
| 12 IRT residual analysis | | | √ | | | | √ | | √ | | √ | √ | √ | | | | | | | | | 5 | Technical report; Field test and simulation analysis |
| 13 Reliability and standard error estimation | | | √ | | | √ | √ | √ | √ | √ | √ | √ | √ | √ | | | | | | | | 5 | Technical report. |
| 14 Cognitive skills and item response time | | √ | | | | √ | | | √ | | √ | | | √ | | | | | | | | 14, 16, 19 | Cognitive lab reports from pilot and field test; Summay response time analysis |
| 15 Cognitive interviews, think-aloud | | √ | | | | √ | | | √ | | √ | | | √ | | | | | | | | 14, 16 | Cognitive lab reports from pilot and field test |
| 16 Decision consistency and accuracy | | | √ | | | | √ | √ | √ | | √ | √ | √ | | √ | | | | | | | 21 | Report from standard setting vendor. |

| # | Item | Ref | Description |
|---|------|-----|-------------|
| 17 | 17 Cut-score standard errors | 21 | Report from standard setting vendor. |
| 18 | 18 Criterion-related validation of on track | Future | Future report from longitudinal operational data. |
| 19 | 19 Educator interviews, focus groups, surveys | 14, 16, 15, 23 | Analysis of pilot and field test survey data; Report of reporting focus groups; surveys of panelists in ALD writing and standard setting. |
| 20 | 20 Criterion-related validation of readiness | Proposed study | Proposed study with Core to College states using external readiness measures |
| 21 | 21 Surveys of postsecondary educators | 12 | Reports of surveys from regional higher ed representatives and state higher ed leads. |
| 22 | 22 Analysis of enrollment, dropout, courses | Future | Future report from longitudinal operational data. |
| 23 | 23 Teacher morale surveys | Future | Future report from longitudinal operational data. |
| 24 | 24 Teacher surveys on changes in students | Future | Future report from longitudinal operational data. |
| 25 | 25 Student morale and aspirations surveys | Future | Future report from longitudinal operational data. |
| 26 | 26 Evaluation of vertical scale | 5 | Pilot report of dimensionality and scaling structure; Field test report of scale construction, stability. |
| 27 | 27 Criterion-related studies re: gain/growth | 5? | Future report from longitudinal operational data. |
| 28 | 28 Follow-up on specific student decisions | Future | Future report from longitudinal operational data. |
| 29 | 29 Sensitivity to instruction | Future | Future report from longitudinal operational data. |
| 30 | 30 Analysis of classroom artifacts | Future | Future report from longitudinal operational data. |
| 31 | 31 Score report utility and clarity | 15 | Focus group reports from scoring contractor. Collective information from state leads and superintendents. |
| 32 | 32 Analysis of report usage rates | 15 | Future report from longitudinal operational data. |
| 33 | 33 Analysis of reliability of aggregate stats | 5, 15 | Psychometirc contractor sets methods, criteria; Reporting vendor carries out analysis |
| 34 | 34 Generalizability studies | 5 | Technical manual. |
| 35 | 35 Item parameter drift | 5, future | Psychometric contractor establishes method; Studies to be conducted with ongoing operational data. |

| # | Item | Reference | Description |
|---|------|-----------|-------------|
| 36 | Audit of UTD and sensitivity review | 14, 16 | Report from pilot and field test item development vendor on item development practices and item sensitiity review methods and execution. |
| 37 | Audit of test accommodations | 6, 19, External studies | Report of accommodations eligibility and delivery from administration vendors. |
| 38 | Differential item functioning | 5 | DIF reports in technical report |
| 39 | Differential predictive validity | Future | Future report from longitudinal operational data. |
| 40 | Invariance of test structure | Future | Future report from longitudinal operational data. |
| 41 | Analysis of group differences | 5, 15 | Test differential functioning (tech report), disaggregated reports by ethnic group, disability, ELL status |
| 42 | Multitrait/multimethod | 5 | Tech report; part of dimensionality study |
| 43 | Scope and sequence curriculum survey | Future | Likely to be carried out by states; |
| 44 | Validation of content clusters | Future | Survey of alignment between teachers' instructional units and interim block assessments. Likely carried out by states. |
| 45 | Analysis of **interim** usage statistics | Future | Future report from longitudinal operational data. |
| 46 | Surveys, interviews, focus groups of (high) users of **interim** assessments | Future | Future report from longitudinal operational data. |
| 47 | Audit of **formative** resources development and implementation | 23, Future | Report from instructional resources vendor of development. Implementation data from operational use. |
| 48 | Analysis of usage stats for **formative** | Future | Future report from longitudinal operational data. |
| 49 | Surveys of Collaborative leadership | Future | Future report from longitudinal operational data. |
| 50 | Educator **formative** assessment surveys | 23 | Report from instructional resources vendor. |
| 51 | **Formative** assessment user surveys | Future | Future report from longitudinal operational data. |
| 52 | Parent, student formative surveys | Future | Future report from longitudinal operational data. |

| # | Task | | | | | | | | | | | | | | | | | | | | | | Status | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53 | 53 Case studies of frequent users | | | | | √ | | | | | | | | | | | | √ | √ | √ | √ | √ | Future | Future report from longitudinal operational data. |
| 54 | 54 Critique of Theory of Action | | | | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | Ongoing | Collect as available |
| 55 | 55 Summary of validity evidence acc. to 7 principles | | | | | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | Res Agenda | Checklist completion--ongoing |
| 56 | 55 Evaluation of local scoringfor interim system | | √ | | | | | | | | | | | | | | | √ | √ | √ | √ | √ | Future | Future report from longitudinal operational data. |

# Checklist Legend

## Summative purposes

| | |
|---|---|
| | **The purposes of the Smarter Balanced *Summative* assessments are to provide valid, reliable and fair information about,** |
| **1** | 1. students' ELA and Mathematics achievement with respect to those CCSS measured by the ELA and Mathematics summative assessments |
| **2** | 2. whether students prior to Grade 11 have demonstrated sufficient academic proficiency in ELA and mathematics to be on track for achieving college readiness |
| **3** | 3. whether Grade 11 students have sufficient academic proficiency in ELA and Mathematics to be ready to take credit-bearing college courses |
| **4** | 4. students' annual progress toward college and career readiness in ELA and Mathematics |
| **5** | 5. how instruction can be improved at the classroom, school, district, and state level |
| **6** | 6. students' ELA and Mathematics proficiencies for Federal accountability purposes and potentially for state and local accountability systems |
| **7** | 7. students' achievement in ELA and Mathematics that is equitable for *all students and subgroups of students* . |

## Interim Purposes

| | |
|---|---|
| | **The purposes of the Smarter Balanced *Interim* assessments are to provide valid, reliable and fair information about,** |
| **1** | 1. student progress toward mastery of the skills measured in ELA and Mathematics by the summative assessment |
| **2** | 2. students' performance at the content cluster level so teachers and administrators can track student progress throughout the year and adjust instruction accordingly |
| **3** | 3. individual and group (e.g., school, district) performance at the claim level in ELA and mathematics to determine whether teaching and learning are on target |
| **4** | 4. student progress toward the mastery of skills measured in ELA and Mathematics *across all subgroups of students* |

## Instructional Resources Purposes

**The purposes of the Smarter Balanced *Formative Assessment Resources* are to provide measurement tools and resources to,**

1. improve teaching and learning
2. monitor student progress throughout the school year
3. help teachers and other educators align instruction, curricula, and assessment
4. help teachers and other educators use the Summative and Interim assessments to improve instruction at the individual student and classroom levels
5. illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning

Appendix N –    2013 Pilot School Districts and Charter Schools

| Status | Response | School Name | Local Education Agency Name |
|---|---|---|---|
| | YES | ADRIAN ELEM. | ADRIAN R-III |
| | YES | ADRIAN SR. HIGH | ADRIAN R-III |
| | YES | ADVANCE ELEM. | ADVANCE R-IV |
| | YES | ADVANCE HIGH | ADVANCE R-IV |
| | NO | ROGERS MIDDLE | AFFTON 101 |
| | YES | ALBANY MIDDLE | ALBANY R-III |
| | YES | VIRGINIA E. GEORGE ELEM. | ALBANY R-III |
| | YES | ALLEN VILLAGE SCHOOL | ALLEN VILLAGE |
| | YES | ALTENBURG ELEM. | ALTENBURG 48 |
| | YES | ALTON ELEM. | ALTON R-IV |
| | YES | APPLETON CITY HIGH | APPLETON CITY R-II |
| | YES | CASS CO. ELEM. | ARCHIE R-V |
| | YES | ASH GROVE ELEM. | ASH GROVE R-IV |
| | NO | ASH GROVE HIGH | ASH GROVE R-IV |
| | YES | AURORA JR. HIGH | AURORA R-VIII |
| NEW | YES | ROBINSON INTERMEDIATE | AURORA R-VIII |
| | YES | AVA MIDDLE | AVA R-I |
| NEW | | Avilla school | Avilla R-13 School District |
| | YES | BALLARD HIGH | BALLARD R-II |
| | YES | BELL CITY ELEM. | BELL CITY R-II |
| | Nonresponsive | BELLEVIEW ELEM. | BELLEVIEW R-III |
| | YES | CAMBRIDGE ELEM. | BELTON 124 |
| NEW | | Hillcrest Elementary | Belton School District #124 |
| | YES | BERNIE ELEM. | BERNIE R-XIII |
| | YES | BERNIE HIGH | BERNIE R-XIII |
| | YES | BEVIER HIGH | BEVIER C-4 |
| NEW | | Blair Oaks Elementary | Blair Oaks R-II |
| NEW | | Blair Oaks Middle | Blair Oaks R-II |
| NEW | YES | BLUE EYE HIGH | BLUE EYE R-V |
| | YES | BLUE SPRINGS HIGH | BLUE SPRINGS R-IV |
| | YES | FRESHMAN CRT. - G. BAKER BLDG. | BLUE SPRINGS R-IV |
| | YES | JOHN NOWLIN ELEM. | BLUE SPRINGS R-IV |
| | YES | BOLIVAR INTERMEDIATE SCH. | BOLIVAR R-1 |
| | YES | BOLIVAR MIDDLE | BOLIVAR R-1 |
| | YES | BOONVILLE HIGH | BOONVILLE R-I |
| NEW | YES | BOWLING GREEN ELEM. | BOWLING GREEN R-I |
| NEW | YES | BOWLING GREEN MIDDLE | BOWLING GREEN R-I |
| | YES | BRECKENRIDGE HIGH | BRECKENRIDGE R-I |
| | YES | BRONAUGH ELEM. | BRONAUGH R-VII |
| | YES | BROOKFIELD HIGH | BROOKFIELD R-III |
| | YES | BROOKSIDE CHARTER SCH. | BROOKSIDE CHARTER SCH. |
| | Nonresponsive | Bunker Elem. | BUNKER R-III |
| | YES | Cainsville High | CAINSVILLE R-I |
| | NO | Cameron Middle | CAMERON R-I |
| xx | Nonresponsive | Turning Point | CAMERON R-I |
| | YES | CANTON ELEM. | CANTON R-V |
| NEW | | Carl Junction Intermediate | Carl Junction R-1 School District |
| NEW | | Carl Junction Intermediate | Carl Junction School District |
| | YES | CARTHAGE JR. HIGH | CARTHAGE R-IX |
| | YES | FAIRVIEW ELEM. | CARTHAGE R-IX |
| | YES | STEADLEY ELEM. | CARTHAGE R-IX |
| | NO | CARTHAGE TECHNICAL CENTER-NOR | CARTHAGE R-IX |
| | YES | CASSVILLE INTERMEDIATE | CASSVILLE R-IV |
| | YES | CENTER MIDDLE | CENTER 58 |

| | | | |
|---|---|---|---|
| | YES | CENTER SR. HIGH | CENTER 58 |
| NEW | | West Elementary | Central R-3 |
| | NO | CENTRALIA INTERMEDIATE | CENTRALIA R-VI |
| | YES | WARREN E. HEARNES ELEM. | CHARLESTON R-I |
| | Nonresponsive | CHARLESSTON HIGH | CHARLESTON R-I |
| | YES | CHILHOWEE ELEM. | CHILHOWEE R-IV |
| | Nonresponsive | CITY GARDEN MONTESSORI SCHOOL | CITY GARDEN MONTESSORI |
| | Nonresponsive | CLARK CO. MIDDLE | CLARK CO. R-I |
| XX | Nonresponsive | CLARKTON ELEM. | CLARKTON C-4 |
| | YES | GLENRIDGE ELEM. | CLAYTON |
| | YES | COLE CO. R-I MIDDLE | COLE CO. R-I |
| | YES | FAIRVIEW ELEM. | COLUMBIA 93 |
| | YES | FREDRICK DOUGLASS HIGH | COLUMBIA 93 |
| | YES | JOHN RIDGEWAY ELEM. | COLUMBIA 93 |
| | YES | MARY PAXTON KEELEY ELEM. | COLUMBIA 93 |
| | YES | NEW HAVEN ELEM. | COLUMBIA 93 |
| | YES | ROBERT E. LEE ELEM. | COLUMBIA 93 |
| | YES | RUSSELL BLVD. ELEM. | COLUMBIA 93 |
| | YES | ULYSSES S. GRANT ELEM. | COLUMBIA 93 |
| | YES | WEST BLVD. ELEM. | COLUMBIA 93 |
| | NO | JUVENILE JUSTICE CTR. | COLUMBIA 93 |
| | NO | MILL CREEK ELEM. | COLUMBIA 93 |
| NEW | YES | COMMUNITY HIGH | COMMUNITY R-VI |
| | YES | CONCORDIA ELEM. | CONCORDIA R-II |
| | Nonresponsive | CONSTRUCTION CAREERS CENTER | CONSTRUCTION CAREERS CENTER |
| | Nonresponsive | COOTER ELEM. | COOTER R-IV |
| | YES | CRANE MIDDLE | CRANE R-III |
| | YES | CROCKER HIGH | CROCKER R-II |
| | Nonresponsive | CRYSTAL CITY ELEM. | CRYSTAL CITY 47 |
| | Nonresponsive | CRYSTAL CITY HIGH | CRYSTAL CITY 47 |
| | YES | DAVIS ELEM. | DAVIS R-XII |
| | Nonresponsive | DELLA LAMB ELEM | DELLA LAMB ELEM |
| | YES | ATHENA ELEM. | DESOTO 73 |
| | YES | T.S. HILL MIDDLE | DEXTER R-XI |
| | NO | ALTERNATIVE RESOURCE CTR. | DIVISION OF YOUTH SERVICES |
| | YES | DIXON ELEM. | DIXON R-I |
| NEW | | Dixon High School | Dixon R-I Schools |
| | Nonresponsive | DORA ELEM. | DORA R-III |
| | NO | PEVELY ELEM. | DUNKLIN R-V |
| | Nonresponsive | HERCULANEUM HIGH SCHOOL | DUNKLIN R-V |
| | YES | EAST BUCHANAN HIGH | EAST BUCHANAN CO. C-1 |
| | YES | A. J. MARTIN ELEM. | EAST PRAIRIE R-II |
| | Nonresponsive | EL DORADO SPRINGS MIDDLE | EL DORADO SPRINGS R-II |
| | Nonresponsive | ELDON UPPER ELEM. | ELDON R-I |
| | YES | EMINENCE ELEM. | EMINENCE R-I |
| | YES | EMINENCE HIGH | EMINENCE R-I |
| | Nonresponsive | EXCELSIOR SPRINGS CAREER CTR. | EXCELSIOR SPRINGS 40 |
| | YES | FAIR GROVE MIDDLE | FAIR GROVE R-X |
| | YES | FAIRFAX ELEM. | FAIRFAX R-III |
| | YES | WASHINGTON-FRANKLIN ELEM. | FARMINGTON R-VII |
| | YES | AIRPORT ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | BERMUDA ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | CROSS KEYS MIDDLE | FERGUSON-FLORISSANT R-II |
| | YES | HOLMAN ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | JOHNSON WABASH ELEM. | FERGUSON-FLORISSANT R-II |

| | | | |
|---|---|---|---|
| | YES | LEE HAMILTON ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | MCCLUER NORTH HIGH | FERGUSON-FLORISSANT R-II |
| | YES | PARKER ROAD ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | WALNUT GROVE ELEM. | FERGUSON-FLORISSANT R-II |
| | YES | FESTUS INTERMEDIATE | FESTUS R-VI |
| | YES | FORDLAND HIGH SCHOOL | FORDLAND R-III |
| | YES | FORDLAND MIDDLE | FORDLAND R-III |
| | YES | FORSYTH HIGH | FORSYTH R-III |
| | YES | FORSYTH MIDDLE | FORSYTH R-III |
| | YES | BUCKNER ELEM. | FORT OSAGE R-I |
| | YES | FIRE PRAIRIE MIDDLE | FORT OSAGE R-I |
| | YES | ANTONIA ELEM. | FOX C-6 |
| | YES | GEORGE GUFFEY ELEM. | FOX C-6 |
| | YES | SECKMAN HIGH SCHOOL | FOX C-6 |
| | YES | FRANCIS HOWELL MIDDLE | FRANCIS HOWELL R-III |
| | YES | HENDERSON ELEM. | FRANCIS HOWELL R-III |
| XX | Nonresponsive | CASTLIO ELEM. | FRANCIS HOWELL R-III |
| | YES | FRANKLIN CO. ELEM. | FRANKLIN CO. R-II |
| NEW | | Fredericktown Intermediate School | Fredericktown R-I School District |
| NEW | | Fredericktown Middle School | Fredericktown R-I School District |
| | Nonresponsive | FRONTIER SCHOOL OF INNOVATION | FRONTIER SCHOOL OF INNOVATION |
| | YES | FOREST PARK ELEM. | FT. ZUMWALT R-II |
| | YES | OSTMANN ELEM. | FT. ZUMWALT R-II |
| | YES | ROCK CREEK ELEM. | FT. ZUMWALT R-II |
| | NO | BARTLEY ELEM. | FULTON 58 |
| | NO | MCINTIRE ELEM. | FULTON 58 |
| | YES | GAINSVILLE HIGH | GAINESVILLE R-V |
| | YES | HERMANN MIDDLE | GASCONADE CO. R-I |
| | Nonresponsive | GATEWAY SCIENCE ACAD/ST LOUIS | GATEWAY SCIENCE ACAD/ST LOUIS |
| | YES | GENESIS SCHOOL INC. | GENESIS SCHOOL INC. |
| | YES | GIDEON ELEM. | GIDEON 37 |
| | Nonresponsive | GILLIAM ELEM. | GILLIAM C-4 |
| | YES | GILMAN CITY ELEM. | GILMAN CITY R-IV |
| | Nonresponsive | GLENWOOD ELEM. | GLENWOOD R-VIII |
| | Nonresponsive | GORDON PARKS ELEM | GORDON PARKS ELEM |
| | YES | MATTHEWS ELEM. | GRAIN VALLEY R-V |
| | YES | GREEN CITY ELEM. | GREEN CITY R-I |
| | YES | GREEN CITY HIGH SCHOOL | GREEN CITY R-I |
| | YES | GREEN FOREST ELEM. | GREEN FOREST R-II |
| | YES | GREEN RIDGE ELEM. | GREEN RIDGE R-VIII |
| NEW | | Hale R-1 | Hale R-1 |
| | YES | HAMILTON ELEM. | HAMILTON R-II |
| | YES | OAKWOOD ELEM. | HANNIBAL 60 |
| | YES | VETERANS ELEM. | HANNIBAL 60 |
| | YES | HARDIN-CENTRAL ELEM. | HARDIN-CENTRAL C-2 |
| | YES | HARDIN-CENTRAL HIGH | HARDIN-CENTRAL C-2 |
| | YES | HARRISONVILLE MIDDLE | HARRISONVILLE R-IX |
| | Nonresponsive | HARTVILLE ELEM. | HARTVILLE R-II |
| | YES | CENTRAL MIDDLE | HAZELWOOD |
| | YES | EAST MIDDLE | HAZELWOOD |
| | YES | JANA ELEM. | HAZELWOOD |
| | YES | LARIMORE ELEM. | HAZELWOOD |
| | YES | TOWNSEND ELEM. | HAZELWOOD |
| | YES | TWILLMAN ELEM. | HAZELWOOD |

| | | | |
|---|---|---|---|
| | YES | WINDSOR HIGH | HENRY CO. R-I |
| | Nonresponsive | WINDSOR ELEM. | HENRY CO. R-I |
| NEW | | SKYLINE HIGHSCHOOL | HICKORY COUNTY R-1 SCHOOLS |
| NEW | | SKYLINE HIGHSCHOOL | HICKORY COUNTY R-1 SCHOOLS |
| | YES | HIGH POINT ELEM. | HIGH POINT R-III |
| | YES | HILLSBORO JR. HIGH | HILLSBORO R-III |
| | YES | HILLSBORO MIDDLE ELEM. | HILLSBORO R-III |
| | YES | HOGAN PREPARATORY ACADEMY | HOGAN PREPARATORY ACADEMY |
| | YES | HOLLISTER Middle School | HOLLISTER R-V |
| | YES | HOWELL VALLEY ELEM. | HOWELL VALLEY R-I |
| | YES | HUDSON ELEM. | HUDSON R-IX |
| | YES | HURLEY ELEM. | HURLEY R-I |
| | YES | IBERIA HIGH | IBERIA R-V |
| | YES | BRYANT ELEM. | INDEPENDENCE 30 |
| | YES | JOHN W. LUFF ELEM. | INDEPENDENCE 30 |
| | YES | SANTA FE TRAIL ELEM. | INDEPENDENCE 30 |
| | YES | SYCAMORE HILLS ELEM. | INDEPENDENCE 30 |
| | YES | THREE TRAILS ELEM. | INDEPENDENCE 30 |
| | YES | WILLIAM SOUTHERN ELEM. | INDEPENDENCE 30 |
| NEW | | George Caleb Bingham Middle School | Independence 30 |
| NEW | | Pioneer Ridge Middle School | Independence 30 |
| NEW | | Spring Branch Elementary | Independence 30 |
| NEW | | Truman High School | Independence 30 |
| NEW | | William Chrisman High School | Independence 30 |
| | YES | VIBURNUM HIGH | IRON CO. C-4 |
| | YES | NORTH ELEM. | JACKSON R-II |
| | YES | ORCHARD DRIVE ELEM. | JACKSON R-II |
| | YES | JASPER HIGH | JASPTER CO. R-V |
| | YES | JEFFERSON HIGH | JEFFERSON C-123 |
| NEW | | Jefferson C-123 | Jefferson C-123 |
| | YES | JEFFERSON CITY HIGH | JEFFERSON CITY |
| | YES | TELEGRAPH INTERMEDIATE | JEFFERSON CO. R-VII |
| NEW | | Jefferson High School | Jefferson County R7 Schools |
| NEW | | Danby Rush Tower Middle School | Jefferson County R-VII School District |
| | YES | ROYAL HEIGHTS ELEM. | JOPLIN SCHOOLS |
| | YES | JUNCTION HILL ELEM. | JUNCTION HILL C-12 |
| | YES | ATTUCKS ELCK. | KANSAS CITY 33 |
| | YES | GLADSTONE ELEM. | KANSAS CITY 33 |
| | YES | HOLLIDAY MONTESSORI | KANSAS CITY 33 |
| | YES | PASEO ACAD. OF PERFORMING ARTS | KANSAS CITY 33 |
| | YES | PHILLIS WHEATLEY ELEM. | KANSAS CITY 33 |
| | YES | TRAILWOODS ELEM. | KANSAS CITY 33 |
| | NO | CONTRACT | KANSAS CITY 33 |
| | NO | EAST HIGH SCHOOL | KANSAS CITY 33 |
| | NO | NORTHEAST HIGH SCHOOL | KANSAS CITY 33 |
| | YES | DOGWOOD ELEM. | KEARNEY R-I |
| | YES | KING CITY ELEM. | KING CITY R-I |
| | NO | KINGSVILLE HIGH | KINGSVILLE R-I |
| | YES | KIRBYVILLE MIDDLE | KIRBYVILLE R-VI |
| | NO | KIRKSVILLE AREA TECH CTR. | KIRKSVILLE R-III |
| | NO | KIRKSVILLE SR. HIGH | KIRKSVILLE R-III |
| | YES | LA PLATA ELEM. | LA PLATA R-II |
| | YES | LA PLATA HIGH | LA PLATA R-II |

| | | | |
|---|---|---|---|
| | YES | EZARD ELEM. | LACLEDE CO. R-I |
| | YES | LADUE MIDDLE | LADUE |
| | YES | SPOEDE ELEM. | LADUE |
| | Nonresponsive | GRANDVIEW ELEM. | LAFAYETTE CO. C-1 |
| | Nonresponsive | LAFAYETTE CO. MIDDLE | LAFAYETTE CO. C-1 |
| | NO | LAKELAND HIGH | LAKELAND R-III |
| | Nonresponsive | LAKELAND ELEM. | LAKELAND R-III |
| | YES | LAMAR MIDDLE | LAMAR R-I |
| | YES | LAQUEY R-V HIGH | LAQUEY R-V |
| | YES | LAREDO ELEM. | LAREDO R-VII |
| | YES | CEDAR CREEK ELEM. | LEE'S SUMMIT R-VII |
| | YES | HAZEL GROVE ELEM. | LEE'S SUMMIT R-VII |
| | YES | MEADOW LANE ELEM. | LEE'S SUMMIT R-VII |
| | YES | RICHARDSON ELEM. | LEE'S SUMMIT R-VII |
| | YES | SUMMIT POINTE ELEM. | LEE'S SUMMIT R-VII |
| | NO | HILLTOP SCHOOL | LEE'S SUMMIT R-VII |
| | NO | LEE'S SUMMIT SR. HIGH | LEE'S SUMMIT R-VII |
| | YES | LEETON ELEM. | LEETON R-X |
| | YES | HIGHLAND ELEM. | LEWIS CO. C-1 |
| | YES | LIBERAL HIGH | LIBERAL R-II |
| | YES | LIBERTY OAKS ELEM. | LIBERTY 53 |
| | Nonresponsive | LIBERTY MIDDLE SCHOOL | LIBERTY 53 |
| | YES | LINCOLN ELEM. | LINCOLN R-II |
| | YES | ROBERT H. SPERRENG MIDDLE | LINDBERGH SCHOOLS |
| | YES | LONE JACK HIGH School | LONE JACK C-6 |
| | YES | LONEDELL ELEM. | LONEDELL R-XIV |
| | YES | MADISON ELEM. | MADISON C-3 |
| | YES | MANSFIELD HIGH | MANSFIELD R-IV |
| | Nonresponsive | MRH ELEMENTARY | MAPLEWOOD-RICHMOND HEIGHTS |
| | Nonresponsive | MARCELINE MIDDLE | MARCELINE R-V |
| | YES | MARION C. EARLY ELEM. | MARION C. EARLY R-V |
| | YES | MARION C. EARLY HIGH | MARION C. EARLY R-V |
| | NO | MARIONVILLE ELEM. | MARIONVILLE R-IX |
| | NO | MARIONVILLE MIDDLE | MARIONVILLE R-IX |
| | Nonresponsive | BUEKER MIDDLE | MARSHALL |
| | Nonresponsive | MARSHALL SR. HIGH | MARSHALL |
| | NO | MARYVILLE HIGH | MARYVILLE R-II |
| | NO | MARYVILLE MIDDLE | MARYVILLE R-II |
| | NO | NORTHWEST TECHNICAL SCH. | MARYVILLE R-II |
| | YES | NOEL ELEM. | MCDONALD CO. R-I |
| | Nonresponsive | BEASLEY ELEM. | MEHLVILLE R-IX |
| | YES | MISSOURI CITY ELEM. | MISSOURI CITY 56 |
| | YES | MOBERLY MIDDLE | MOBERLY |
| | YES | NORTH CENTRAL REGIONAL | MOBERLY |
| NEW | | Monett R-1 School District | Monett R-1 Schools |
| | YES | CALIFORNIA ELEM. | MONITEAU CO. R-I |
| | YES | MONROE CITY MIDDLE | MONROE CITY R-I |
| | Nonresponsive | MONTROSE ELEM. | MONTROSE R-XIV |
| NEW | | Morgan County R-1 Elementary | Morgan County R-1 School district |
| | YES | MOUNTAIN GROVE ELEM. | MOUNTAIN GROVE R-III |
| | YES | MT. VERNON MIDDLE | MT. VERNON R-V |
| | NO | NAYLOR ELEM. | NAYLOR R-II |
| | YES | NEELYVILLE HIGH | NEELYVILLE R-IV |
| | YES | CENTRAL ELEM. | NEOSHO R-V |
| | YES | GEORGE WASHINGTON CARVER ELE | NEOSHO R-V |

| | | | |
|---|---|---|---|
| | YES | NEVADA HIGH | NEVADA R-V |
| | YES | TRUMAN ELEM. | NEVADA R-V |
| | YES | NEW BLOOMFIELD HIGH | NEW BLOOMFIELD R-III |
| | YES | NEW YORK ELEM. | NEW YORK R-IV |
| | YES | NEWBURG ELEM. | NEWBURG R-II |
| | YES | CENTURY ELEM. | NIXA R-II |
| | YES | NICHOLAS A. INMAN INTERMEDIATE | NIXA R-II |
| | YES | WILLIAMSBURG ELEM. | NORTH CALLAWAY CO. R-I |
| | NO | NORTH DAVIESS ELEM. | NORTH DAVIESS R-III |
| | YES | MEADOWBROOK ELEM. | NORTH KANSAS CITY 74 |
| | NO | MAPLE PARK MIDDLE | NORTH KANSAS CITY 74 |
| | YES | NORTH PLATTE HIGH | NORTH PLATTE CO. R-I |
| | YES | NORTH PLATTE INTERMEDIATE | NORTH PLATTE CO. R-I |
| | YES | NORTH WOOD ELEM. | NORTH WOOD R-IV |
| | YES | NORTHEAST NODAWAY ELEMENTARY | NORTHEAST NODAWAY CO. R-V |
| | YES | BRENNAN WOODS ELEM. | NORTHWEST R-I |
| | YES | ODESSA HIGH | ODESSA R-VII |
| | NO | ORCHARD FARM MIDDLE | ORCHARD FARM R-V |
| | YES | WEST ELEM. | OZARK R-VI |
| | YES | HAWTHORN ELEM. | PARK HILL |
| | YES | BARRETTS ELEM. | PARKWAY C-2 |
| | YES | BELLERIVE ELEM. | PARKWAY C-2 |
| | YES | HANNA WOODS ELEM. | PARKWAY C-2 |
| | YES | SHENANDOAH VALLEY ELEM. | PARKWAY C-2 |
| | YES | SOUTH MIDDLE | PARKWAY C-2 |
| | YES | PATTONSBURG HIGH | PATTONSBURG R-II |
| | YES | ROSE ACRES ELEM. | PATTONVILLE R-III |
| | YES | PETTIS CO. ELEM. | PETTIS CO. R-XII |
| | YES | CLOPTON ELEM. | PIKE CO. R-III |
| | Nonresponsive | PIKE/LINCOLN TECHNICAL CTR. | PIKE CO. R-III |
| | YES | PLATO HIGH | PLATO R-V |
| | NO | BARRY SCH. | PLATTE CO. R-III |
| | NO | DONALD D. SIEGRIST ELEM. | PLATTE CO. R-III |
| NEW | | Polo R-VII High School | Polo R-VII School District |
| NEW | | Polo R-VII School District | Polo R-VII School District |
| | YES | LAKE ROAD ELEM. | POPLAR BLUFF R-I |
| NEW | | John Evans Middle School | Potosi R-III School Distirct |
| NEW | | Potosi Elementary School | Potosi R-III School Distirct |
| NEW | | Potosi High School | Potosi R-III School Distirct |
| NEW | | Trojan Intermediate School | Potosi R-III School Distirct |
| | YES | PRAIRIE HOME ELEM. | PRAIRIE HOME R-V |
| | YES | PURDY MIDDLE | PURDY R-II |
| | YES | PUXICO ELEM. | PUXICO R-VIII |
| | YES | PUXICO JR. HIGH | PUXICO R-VIII |
| | YES | CENTER ELEM. | RALLS CO. R-II |
| | Nonresponsive | MARK TWAIN SR. HIGH | RALLS CO. R-II |
| | YES | BLUE RIDGE ELEM. | RAYTOWN C-2 |
| | YES | FLEETRIDGE ELEM. | RAYTOWN C-2 |
| | YES | REEDS SPRING INTERMEDIATE | REEDS SPRING R-IV |
| | YES | REEDS SPRING MIDDLE | REEDS SPRING R-IV |
| | YES | LYON ELEMENTARY | REPUBLIC R-III |
| | YES | PRICE ELEMENTARY | REPUBLIC R-III |
| | YES | REPUBLIC HIGH | REPUBLIC R-III |
| | YES | REPUBLIC MIDDLE | REPUBLIC R-III |
| | Nonresponsive | RICH HILL HIGH | RICH HILL R-IV |

| | | | |
|---|---|---|---|
| | Nonresponsive | RICHARDS ELEM. | RICHARDS R-V |
| | YES | RICHLAND ELEM. | RICHLAND R-IV |
| | YES | SUNRISE ELEM. | RICHMOND R-XVI |
| NEW | | Ripley County R-III School District | Ripley County R-III School District |
| | Nonresponsive | WYLAND ELEM. | RITENOUR |
| | YES | CRESTVIEW MIDDLE | ROCKWOOD R-VI |
| | YES | RIDGE MEADOWS ELEM. | ROCKWOOD R-VI |
| | YES | WESTRIDGE ELEM. | ROCKWOOD R-VI |
| | Nonresponsive | LAFAYETTE SR. HIGH | ROCKWOOD R-VI |
| | YES | ROLLA MIDDLE | ROLLA 31 |
| NEW | | Salisbury High School | Salisbury R-IV |
| | YES | SANTA FE ELEM. | SANTA FE R-X |
| | NO | SANTA FE HIGH | SANTA FE R-X |
| | YES | WILDWOOD ELEM. | SARCOXIE R-II |
| | YES | MINNIE CLINE ELEM. | SAVANNAH R-III |
| | YES | SAVANNAH MIDDLE | SAVANNAH R-III |
| | Nonresponsive | JOHN GLENN ELEM. | SAVANNAH R-III |
| | Nonresponsive | OSAGE MIDDLE | SCHOOL OF THE OSAGE |
| | YES | SCOTT CITY MIDDLE | SCOTT CITY R-I |
| | NO | THOMAS W. KELLY HIGH | SCOTT CO. R-IV |
| | YES | SCUOLA VITA NUOVA CHARTER | SCUOLA VITA NUOVA |
| | YES | HORNERSVILLE MIDDLE | SENATH-HORNERSVILLE C-8 |
| | Nonresponsive | SENATH-HORNERSVILLE SR. HIGH | SENATH-HORNERSVILLE C-8 |
| | Nonresponsive | SENECA INTERMEDIATE SCHOOL | SENECA R-VII |
| | YES | SEYMOUR HIGH | SEYMOUR R-II |
| | YES | SHELBINA ELEM. | SHELBY CO. R-IV |
| NEW | | Clarence Elementary | Shelby County R-IV School District |
| NEW | YES | SHERWOOD HIGH SCHOOL | SHERWOOD CASS R-VIII |
| | YES | 7TH AND 8TH GRADE CTR. | SIKESTON R-6 |
| NEW | | Silex High School | Silex RI School District |
| | NO | SLATER HIGH | SLATER |
| | YES | SMITHTON ELEM. | SMITHTON R-VI |
| | YES | SMITHVILLE UPPER ELEM. | SMITHVILLE R-II |
| NEW | | South Callaway R-II Elementary | South Callaway R-II School District |
| NEW | | South Callaway R-II High School | South Callaway R-II School District |
| NEW | | South Callaway R-II Middle School | South Callaway R-II School District |
| | NO | SOUTH IRON ELEM. | SOUTH IRON CO. R-I |
| | Nonresponsive | SOUTH PEMISCOT ELEM. | SOUTH PEMISCOT CO. R-V |
| | Nonresponsive | SOUTHERN BOONE MIDDLE | SOUTHERN BOONE CO. R-I |
| | YES | SOUTHWEST LIVINGSTON CO R-1 EL | SOUTHWEST LIVINGSTON CO. R-I |
| | Nonresponsive | EXTERNAL SITES | SPECL. SCH. DST. ST. LOUIS CO. |
| | Nonresponsive | SPICKARD ELEM. | SPICKARD R-II |
| | YES | SPOKANE MIDDLE | SPOKANE R-VII |
| | YES | SPRING BLUFF ELEM. | SPRING BLUFF R-XV |
| | YES | BINGHAM ELEM. | SPRINGFIELD R-XII |
| | YES | CAMPBELL ELEM. | SPRINGFIELD R-XII |
| NEW | YES | HILLCREST HIGH | SPRINGFIELD R-XII |
| | YES | PLEASANT VIEW MIDDLE | SPRINGFIELD R-XII |
| | NO | DELAWARE ELEM. | SPRINGFIELD R-XII |
| | NO | HOLLAND ELEM. | SPRINGFIELD R-XII |
| | NO | MCGREGOR ELEM. | SPRINGFIELD R-XII |
| | NO | WILLIAMS ELEM. | SPRINGFIELD R-XII |

| | | | |
|---|---|---|---|
| | NO | YORK ELEM. | SPRINGFIELD R-XII |
| | YES | BODE MIDDLE | ST. JOSEPH |
| | YES | EDISON ELEM. | ST. JOSEPH |
| | YES | HALL ELEM. | ST. JOSEPH |
| | YES | MARK TWAIN ELEM. | ST. JOSEPH |
| | YES | ROBIDOUX MIDDLE | ST. JOSEPH |
| | YES | BUDER ELEM. | ST. LOUIS CITY |
| | YES | DUNBAR AND BR. | ST. LOUIS CITY |
| | YES | EARL NANCE, SR. ELEM. | ST. LOUIS CITY |
| | YES | FARRAGUT ELEM. | ST. LOUIS CITY |
| | YES | JEFFERSON ELEM. | ST. LOUIS CITY |
| | YES | LACLEDE ELEM. | ST. LOUIS CITY |
| | YES | MALLINCKRODT A.B.I. ELEM. | ST. LOUIS CITY |
| | YES | PEABODY ELEM. | ST. LOUIS CITY |
| | YES | SHENANDOAH ELEM. | ST. LOUIS CITY |
| | YES | SIGEL ELEM. COMM. ED. CTR. | ST. LOUIS CITY |
| | YES | SUMNER HIGH | ST. LOUIS CITY |
| | YES | WOODWARD ELEM. | ST. LOUIS CITY |
| | YES | YEATMAN-LIDDELL MIDDLE SCHOOL | ST. LOUIS CITY |
| | NO | SOLDAN INTERNATIONAL STUDIES | ST. LOUIS CITY |
| NEW | NO | WASHINGTON MONTESSORI | ST. LOUIS CITY |
| | Nonresponsive | MASON ELEM. | ST. LOUIS CITY |
| | Nonresponsive | WALBRIDGE ELEM. COMMUNITY ED | ST. LOUIS CITY |
| | YES | STOCKTON ELEM. | STOCKTON R-I |
| | YES | STRASBURG ELEM. | STRASBURG C-3 |
| | YES | STURGEON MIDDLE | STURGEON R-V |
| | YES | SWEET SPRINGS ELEM. | SWEET SPRINGS R-VII |
| | NO | SWEET SPRINGS HIGH | SWEET SPRINGS R-VII |
| | YES | THORNFIELD ELEM. | THORNFIELD R-I |
| | Nonresponsive | TINA-AVALON ELEM. | TINA-AVALON R-II |
| | YES | TRENTON MIDDLE | TRENTON R-IX |
| | YES | TRI-COUNTY ELEM. | TRI-COUNTY R-VII |
| | YES | HAWK POINT ELEM. | TROY R-III |
| | YES | UNION STAR HIGH | UNION STAR R-II |
| | NO | UNIVERSITY ACADEMY-LOWER | UNIVERISTY ACADEMY |
| | YES | BARBARA JORDAN ELEM. | UNIVERSITY CITY |
| | YES | JACKSON PARK ELEM. | UNIVERSITY CITY |
| NEW | | Valley Park High School | Valley Park School District |
| | YES | VAN BUREN HIGH | VAN BUREN R-I |
| | Nonresponsive | VAN-FAR ELEM. | VAN-FAR R-I |
| | YES | WARRENSBURG MIDDLE | WARRENSBURG R-VI |
| NEW | | Carterville Elementary | Webb City R-7 School District |
| NEW | | Eugene Field Elementary | Webb City R-7 School District |
| NEW | | Harry S Truman | Webb City R-7 School District |
| NEW | | Mark Twain Elementary | Webb City R-7 School District |
| NEW | | Webb City High School | Webb City R-7 School District |
| NEW | | Webb City Junior High | Webb City R-7 School District |
| NEW | | Webb City Middle School | Webb City R-7 School District |
| | YES | HUDSON ELEM. | WEBSTER GROVES |
| | NO | AVERY ELEM. | WEBSTER GROVES |
| | YES | WELLSVILLE ELEM. | WELLSVILLE MIDDLETOWN R-I |
| | YES | WENTZVILLE SOUTH MIDDLE | WENTZVILLE R-IV |
| | YES | SOUTH FORK ELEM. | WEST PLAINS R-VII |

| | | | |
|---|---|---|---|
| | YES | WEST PLAINS ELEM. | WEST PLAINS R-VII |
| | YES | WHEATLAND HIGH | WHEATLAND R-II |
| | NO | MO SCHOOL FOR THE DEAF | Wheeler Middle |
| NEW | | East Elementary | Willard R-II |
| NEW | | North Elementary | Willard R-II |
| NEW | | Orchard Hills Elementary | Willard R-II |
| NEW | | South Elementary | Willard R-II |
| NEW | | Willard Intermediate Schools | Willard R-II |
| NEW | | Willard Middle School | Willard R-II |
| | YES | WILLOW SPRINGS MIDDLE | WILLOW SPRINGS R-IV |
| | YES | ZALMA ELEM. | ZALMA R-V |

Appendix O – 2014 Field Test School Districts and Charter Schools

| Date | TestName | DISTRICTNAME | Opportunity | TotalStudent | TotalStudentStarte | TotalStudentComp | PercentStarted | PercentCompleted |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-ELA-PT-A New | ADAIR CO. R-I | 1 | 23 | 21 | 21 | 91.30% | 91.30% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | ADAIR CO. R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Cell Phor | ADAIR CO. R-I | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | ADAIR CO. R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | ADAIR CO. R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | ADAIR CO. R-I | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | ADAIR CO. R-I | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | ADAIR CO. R-I | 1 | 23 | 22 | 17 | 95.65% | 73.91% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | ADAIR CO. R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | ADVANCE R-IV | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | ADVANCE R-IV | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | ADVANCE R-IV | 1 | 29 | 28 | 28 | 96.55% | 96.55% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | ADVANCE R-IV | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ADVANCE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Deserts | AFFTON 101 | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | ELA-PT-Deserts-A | AFFTON 101 | 1 | 180 | 168 | 160 | 93.33% | 88.89% |
| 5/19/2014 2:00 | ELA-PT-Renewable | AFFTON 101 | 1 | 181 | 174 | 173 | 96.13% | 95.58% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | AFFTON 101 | 1 | 45 | 43 | 23 | 95.56% | 51.11% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | AFFTON 101 | 1 | 46 | 44 | 21 | 95.65% | 45.65% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | AFFTON 101 | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | AFFTON 101 | 1 | 45 | 39 | 26 | 86.67% | 57.78% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | AFFTON 101 | 1 | 44 | 42 | 29 | 95.45% | 65.91% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | AFFTON 101 | 1 | 36 | 33 | 33 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | AFFTON 101 | 1 | 36 | 35 | 34 | 97.22% | 94.44% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | AFFTON 101 | 1 | 36 | 35 | 35 | 97.22% | 97.22% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | AFFTON 101 | 1 | 36 | 34 | 34 | 94.44% | 94.44% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | AFFTON 101 | 1 | 37 | 34 | 33 | 91.89% | 89.19% |
| 5/19/2014 2:00 | Math-PT-South Po | ALBANY R-III | 1 | 35 | 32 | 32 | 91.43% | 91.43% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ALBANY R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ALBANY R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ALBANY R-III | 1 | 8 | 7 | 7 | 87.50% | 87.50% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ALBANY R-III | 1 | 8 | 6 | 6 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | ALBANY R-III | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | APPLETON CITY R- | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | APPLETON CITY R- | 1 | 14 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | APPLETON CITY R- | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | APPLETON CITY R- | 1 | 7 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G05-ELA-No APPLETON CITY R- | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G05-ELA-No APPLETON CITY R- | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G05-ELA-No APPLETON CITY R- | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 ELA-PT-Archeologi ARCADIA VALLEY F | 1 | 66 | 59 | 59 | 89.39% | 89.39% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ARCADIA VALLEY F | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ARCADIA VALLEY F | 1 | 14 | 12 | 12 | 85.71% | 85.71% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ARCADIA VALLEY F | 1 | 13 | 11 | 11 | 84.62% | 84.62% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ARCADIA VALLEY F | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ARCADIA VALLEY F | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 ELA-PT-Archeologi ATLANTA C-3 | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ATLANTA C-3 | 1 | 16 | 16 | 15 | 100.00% | 93.75% |
| 5/19/2014 2:00 SBAC-G08-ELA-No ATLANTA C-3 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Uncommo AVILLA R-XIII | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 SBAC-G04-ELA-No AVILLA R-XIII | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No AVILLA R-XIII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No AVILLA R-XIII | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No AVILLA R-XIII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No AVILLA R-XIII | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 ELA-PT-Technolog BAKERSFIELD R-IV | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Donuts BAKERSFIELD R-IV | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BAKERSFIELD R-IV | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G07-Math-N BAKERSFIELD R-IV | 1 | 24 | 22 | 22 | 91.67% | 91.67% |
| 5/19/2014 2:00 HS-Math-PT-Great BAYLESS | 1 | 125 | 86 | 77 | 68.80% | 61.60% |
| 5/19/2014 2:00 SBAC-HS-Math-No BAYLESS | 1 | 125 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 ELA-PT-Land Form BELL CITY R-II | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-Math-PT-Great BELL CITY R-II | 1 | 16 | 14 | 14 | 87.50% | 87.50% |
| 5/19/2014 2:00 SBAC-G03-ELA-No BELL CITY R-II | 1 | 23 | 20 | 20 | 86.96% | 86.96% |
| 5/19/2014 2:00 SBAC-HS-Math-No BELL CITY R-II | 1 | 16 | 14 | 13 | 87.50% | 81.25% |
| 5/19/2014 2:00 ELA-PT-Renewable BELTON 124 | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 ELA-PT-Renewable BELTON 124 | 1 | 343 | 327 | 324 | 95.34% | 94.46% |
| 5/19/2014 2:00 Math-PT-Commun BELTON 124 | 1 | 64 | 60 | 60 | 93.75% | 93.75% |
| 5/19/2014 2:00 Math-PT-Order Fo BELTON 124 | 1 | 101 | 88 | 88 | 87.13% | 87.13% |
| 5/19/2014 2:00 Math-PT-Turtle Ha BELTON 124 | 1 | 67 | 57 | 57 | 85.07% | 85.07% |
| 5/19/2014 2:00 SBAC-G04-Math-N BELTON 124 | 1 | 47 | 43 | 42 | 91.49% | 89.36% |
| 5/19/2014 2:00 SBAC-G04-Math-N BELTON 124 | 1 | 46 | 41 | 41 | 89.13% | 89.13% |
| 5/19/2014 2:00 SBAC-G04-Math-N BELTON 124 | 1 | 47 | 44 | 44 | 93.62% | 93.62% |
| 5/19/2014 2:00 SBAC-G04-Math-N BELTON 124 | 1 | 46 | 43 | 43 | 93.48% | 93.48% |
| 5/19/2014 2:00 SBAC-G04-Math-N BELTON 124 | 1 | 46 | 45 | 45 | 97.83% | 97.83% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BELTON 124 | 1 | 86 | 84 | 82 | 97.67% | 95.35% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BELTON 124 | 1 | 86 | 85 | 83 | 98.84% | 96.51% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BELTON 124 | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BELTON 124 | 1 | 86 | 82 | 82 | 95.35% | 95.35% |
| 5/19/2014 2:00 SBAC-G07-ELA-No BELTON 124 | 1 | 85 | 83 | 81 | 97.65% | 95.29% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great BERNIE R-XIII | 1 | 31 | 19 | 19 | 61.29% | 61.29% |
| 5/19/2014 2:00 | SBAC-HS-Math-No BERNIE R-XIII | 1 | 31 | 19 | 19 | 61.29% | 61.29% |
| 5/19/2014 2:00 | HS-Math-PT-Great BEVIER C-4 | 1 | 18 | 14 | 14 | 77.78% | 77.78% |
| 5/19/2014 2:00 | SBAC-HS-Math-No BEVIER C-4 | 1 | 18 | 16 | 16 | 88.89% | 88.89% |
| 5/19/2014 2:00 | HS-ELA-PT-A New BISMARCK R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New BISMARCK R-V | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great BISMARCK R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl BISMARCK R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl BISMARCK R-V | 1 | 11 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl BISMARCK R-V | 1 | 11 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No BISMARCK R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great BLUE EYE R-V | 1 | 53 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No BLUE EYE R-V | 1 | 53 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W BLUE SPRINGS R-IV | 1 | 72 | 68 | 67 | 94.44% | 93.06% |
| 5/19/2014 2:00 | ELA-PT-Deserts-A BLUE SPRINGS R-IV | 1 | 83 | 64 | 63 | 77.11% | 75.90% |
| 5/19/2014 2:00 | ELA-PT-Technology BLUE SPRINGS R-IV | 1 | 238 | 206 | 204 | 86.55% | 85.71% |
| 5/19/2014 2:00 | ELA-PT-The Americ BLUE SPRINGS R-IV | 1 | 80 | 76 | 75 | 95.00% | 93.75% |
| 5/19/2014 2:00 | ELA-PT-Trees-A BLUE SPRINGS R-IV | 1 | 72 | 49 | 48 | 68.06% | 66.67% |
| 5/19/2014 2:00 | HS-Math-PT-Great BLUE SPRINGS R-IV | 1 | 601 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo BLUE SPRINGS R-IV | 1 | 147 | 131 | 130 | 89.12% | 88.44% |
| 5/19/2014 2:00 | Math-PT-Sandbox- BLUE SPRINGS R-IV | 1 | 85 | 76 | 76 | 89.41% | 89.41% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor BLUE SPRINGS R-IV | 1 | 23 | 15 | 14 | 65.22% | 60.87% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor BLUE SPRINGS R-IV | 1 | 25 | 18 | 18 | 72.00% | 72.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor BLUE SPRINGS R-IV | 1 | 24 | 17 | 17 | 70.83% | 70.83% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor BLUE SPRINGS R-IV | 1 | 72 | 68 | 9 | 94.44% | 12.50% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor BLUE SPRINGS R-IV | 1 | 16 | 15 | 12 | 93.75% | 75.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor BLUE SPRINGS R-IV | 1 | 17 | 14 | 14 | 82.35% | 82.35% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor BLUE SPRINGS R-IV | 1 | 16 | 14 | 13 | 87.50% | 81.25% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor BLUE SPRINGS R-IV | 1 | 17 | 11 | 10 | 64.71% | 58.82% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor BLUE SPRINGS R-IV | 1 | 17 | 11 | 9 | 64.71% | 52.94% |
| 5/19/2014 2:00 | SBAC-G04-Math-N BLUE SPRINGS R-IV | 1 | 30 | 28 | 27 | 93.33% | 90.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N BLUE SPRINGS R-IV | 1 | 29 | 25 | 23 | 86.21% | 79.31% |
| 5/19/2014 2:00 | SBAC-G04-Math-N BLUE SPRINGS R-IV | 1 | 30 | 27 | 26 | 90.00% | 86.67% |
| 5/19/2014 2:00 | SBAC-G04-Math-N BLUE SPRINGS R-IV | 1 | 29 | 25 | 24 | 86.21% | 82.76% |
| 5/19/2014 2:00 | SBAC-G04-Math-N BLUE SPRINGS R-IV | 1 | 29 | 28 | 26 | 96.55% | 89.66% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor BLUE SPRINGS R-IV | 1 | 20 | 20 | 16 | 100.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor BLUE SPRINGS R-IV | 1 | 20 | 17 | 13 | 85.00% | 65.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor BLUE SPRINGS R-IV | 1 | 20 | 19 | 18 | 95.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor BLUE SPRINGS R-IV | 1 | 20 | 20 | 16 | 100.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BLUE SPRINGS R-IV | 1 | 22 | 18 | 18 | 81.82% | 81.82% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BLUE SPRINGS R-IV | 1 | 21 | 18 | 18 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BLUE SPRINGS R-IV | 1 | 21 | 19 | 18 | 90.48% | 85.71% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BLUE SPRINGS R-IV | 1 | 21 | 19 | 19 | 90.48% | 90.48% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G06-ELA-No BLUE SPRINGS R-IV | 1 | 59 | 48 | 44 | 81.36% | 74.58% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BLUE SPRINGS R-IV | 1 | 60 | 49 | 44 | 81.67% | 73.33% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BLUE SPRINGS R-IV | 1 | 59 | 52 | 48 | 88.14% | 81.36% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BLUE SPRINGS R-IV | 1 | 60 | 56 | 53 | 93.33% | 88.33% |
| 5/19/2014 2:00 | SBAC-HS-Math-No BLUE SPRINGS R-IV | 1 | 601 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an BOLIVAR R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A BOLIVAR R-I | 1 | 185 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh BOLIVAR R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BOLIVAR R-I | 1 | 1 | 1 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N BOLIVAR R-I | 1 | 43 | 43 | 43 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N BOLIVAR R-I | 1 | 39 | 39 | 39 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N BOLIVAR R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N BOLIVAR R-I | 1 | 50 | 42 | 42 | 84.00% | 84.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N BOLIVAR R-I | 1 | 53 | 44 | 44 | 83.02% | 83.02% |
| 5/19/2014 2:00 | HS-ELA-PT-A New BOWLING GREEN I | 1 | 73 | 73 | 73 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New BOWLING GREEN I | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox BOWLING GREEN I | 1 | 80 | 75 | 74 | 93.75% | 92.50% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BOWLING GREEN I | 1 | 20 | 19 | 19 | 95.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BOWLING GREEN I | 1 | 20 | 19 | 19 | 95.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BOWLING GREEN I | 1 | 20 | 20 | 20 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N BOWLING GREEN I | 1 | 20 | 19 | 19 | 95.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non BOWLING GREEN I | 1 | 73 | 73 | 73 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non BOWLING GREEN I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non BOWLING GREEN I | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts BRADLEYVILLE R-I | 1 | 12 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N BRADLEYVILLE R-I | 1 | 12 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Importanc BRANSON R-IV | 1 | 180 | 176 | 176 | 97.78% | 97.78% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BRANSON R-IV | 1 | 45 | 44 | 43 | 97.78% | 95.56% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BRANSON R-IV | 1 | 45 | 44 | 43 | 97.78% | 95.56% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BRANSON R-IV | 1 | 45 | 45 | 45 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No BRANSON R-IV | 1 | 45 | 44 | 44 | 97.78% | 97.78% |
| 5/19/2014 2:00 | ELA-PT-Trees-A BRAYMER C-4 | 1 | 20 | 19 | 19 | 95.00% | 95.00% |
| 5/19/2014 2:00 | Math-PT-Baseball- BRAYMER C-4 | 1 | 23 | 21 | 21 | 91.30% | 91.30% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N BRAYMER C-4 | 1 | 9 | 8 | 8 | 88.89% | 88.89% |
| 5/19/2014 2:00 | SBAC-G08-Math-N BRAYMER C-4 | 1 | 10 | 9 | 9 | 90.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N BRAYMER C-4 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New BRECKENRIDGE R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great BRECKENRIDGE R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | BRECKENRIDGE R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | BRECKENRIDGE R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | BROOKFIELD R-III | 1 | 73 | 71 | 71 | 97.26% | 97.26% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | BROOKFIELD R-III | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | BROOKFIELD R-III | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | BROOKFIELD R-III | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | BROOKFIELD R-III | 1 | 19 | 17 | 17 | 89.47% | 89.47% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | BRUNSWICK R-II | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | BRUNSWICK R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | BRUNSWICK R-II | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BRUNSWICK R-II | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | BUCHANAN CO. R- | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | BUCHANAN CO. R- | 1 | 17 | 11 | 11 | 64.71% | 64.71% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | BUTLER R-V | 1 | 85 | 22 | 1 | 25.88% | 1.18% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | BUTLER R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | BUTLER R-V | 1 | 105 | 2 | 1 | 1.90% | 0.95% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 21 | 20 | 13 | 95.24% | 61.90% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 21 | 20 | 8 | 95.24% | 38.10% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 2 | 1 | 0 | 50.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 21 | 19 | 12 | 90.48% | 57.14% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 21 | 18 | 11 | 85.71% | 52.38% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | BUTLER R-V | 1 | 21 | 20 | 7 | 95.24% | 33.33% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | BUTLER R-V | 1 | 17 | 11 | 0 | 64.71% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | BUTLER R-V | 1 | 17 | 13 | 2 | 76.47% | 11.76% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | BUTLER R-V | 1 | 17 | 7 | 0 | 41.18% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | BUTLER R-V | 1 | 17 | 5 | 0 | 29.41% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | BUTLER R-V | 1 | 17 | 3 | 0 | 17.65% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | CAINSVILLE R-I | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAINSVILLE R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAINSVILLE R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAINSVILLE R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAINSVILLE R-I | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | CALLAO C-8 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | CALLAO C-8 | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | CALLAO C-8 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 2 | 2 | 1 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 2 | 2 | 1 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | CALLAO C-8 | 1 | 1 | 1 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CALLAO C-8 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | CAMDENTON R-III | 1 | 250 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | Math-PT-Making S | CAMDENTON R-III | 1 | 250 | 230 | 230 | 92.00% | 92.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | CAMDENTON R-III | 1 | 250 | 1 | 1 | 0.40% | 0.40% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | CAMDENTON R-III | 1 | 140 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | CAMDENTON R-III | 1 | 110 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | CAMERON R-I | 1 | 125 | 96 | 93 | 76.80% | 74.40% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAMERON R-I | 1 | 62 | 46 | 32 | 74.19% | 51.61% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | CAMERON R-I | 1 | 63 | 50 | 34 | 79.37% | 53.97% |
| 5/19/2014 2:00 | Math-PT-Commun | CANTON R-V | 1 | 35 | 31 | 31 | 88.57% | 88.57% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CANTON R-V | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CANTON R-V | 1 | 7 | 5 | 5 | 71.43% | 71.43% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CANTON R-V | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CANTON R-V | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CANTON R-V | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | ELA-PT-Animals W | CAPE GIRARDEAU | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | CAPE GIRARDEAU | 1 | 61 | 54 | 54 | 88.52% | 88.52% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | CAPE GIRARDEAU | 1 | 20 | 19 | 18 | 95.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | CAPE GIRARDEAU | 1 | 20 | 20 | 19 | 100.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | CAPE GIRARDEAU | 1 | 21 | 20 | 18 | 95.24% | 85.71% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | CAPE GIRARDEAU | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | CARL JUNCTION R- | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | CARL JUNCTION R- | 1 | 263 | 246 | 246 | 93.54% | 93.54% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 86 | 18 | 18 | 20.93% | 20.93% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 91 | 21 | 21 | 23.08% | 23.08% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CARL JUNCTION R- | 1 | 52 | 17 | 17 | 32.69% | 32.69% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | CARROLLTON R-VI | 1 | 66 | 65 | 64 | 98.48% | 96.97% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | CARROLLTON R-VI | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | CARROLLTON R-VI | 1 | 63 | 51 | 51 | 80.95% | 80.95% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 13 | 11 | 11 | 84.62% | 84.62% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | CARROLLTON R-VI | 1 | 13 | 7 | 7 | 53.85% | 53.85% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CARROLLTON R-VI | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CARROLLTON R-VI | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CARROLLTON R-VI | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CARROLLTON R-VI | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CARROLLTON R-VI | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | CARTHAGE R-IX | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | CARTHAGE R-IX | 1 | 32 | 28 | 28 | 87.50% | 87.50% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | Math-PT-Animal Ju CARTHAGE R-IX | 1 | 93 | 87 | 87 | 93.55% | 93.55% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No CARTHAGE R-IX | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No CARTHAGE R-IX | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No CARTHAGE R-IX | 1 | 11 | 11 | 11 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No CARTHAGE R-IX | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CARTHAGE R-IX | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CARTHAGE R-IX | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CARTHAGE R-IX | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CARTHAGE R-IX | 1 | 19 | 16 | 16 | 84.21% | 84.21% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CARTHAGE R-IX | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | Math-PT-Cell Phor CASSVILLE R-IV | 1 | 157 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N CASSVILLE R-IV | 1 | 30 | 30 | 11 | 100.00% | 36.67% |
| 5/19/2014 2:00 | SBAC-G07-Math-N CASSVILLE R-IV | 1 | 32 | 32 | 9 | 100.00% | 28.13% |
| 5/19/2014 2:00 | SBAC-G07-Math-N CASSVILLE R-IV | 1 | 48 | 33 | 7 | 68.75% | 14.58% |
| 5/19/2014 2:00 | SBAC-G07-Math-N CASSVILLE R-IV | 1 | 47 | 35 | 12 | 74.47% | 25.53% |
| 5/19/2014 2:00 | Math-PT-Order Fo CENTER 58 | 1 | 96 | 41 | 40 | 42.71% | 41.67% |
| 5/19/2014 2:00 | Math-PT-Talent Sh CENTER 58 | 1 | 101 | 36 | 36 | 35.64% | 35.64% |
| 5/19/2014 2:00 | SBAC-G03-Math-N CENTER 58 | 1 | 96 | 41 | 35 | 42.71% | 36.46% |
| 5/19/2014 2:00 | SBAC-G05-Math-N CENTER 58 | 1 | 25 | 8 | 8 | 32.00% | 32.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N CENTER 58 | 1 | 25 | 9 | 8 | 36.00% | 32.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N CENTER 58 | 1 | 25 | 10 | 9 | 40.00% | 36.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N CENTER 58 | 1 | 26 | 9 | 8 | 34.62% | 30.77% |
| 5/19/2014 2:00 | HS-ELA-PT-A New CHAFFEE R-II | 1 | 31 | 26 | 26 | 83.87% | 83.87% |
| 5/19/2014 2:00 | HS-ELA-PT-A New CHAFFEE R-II | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha CHAFFEE R-II | 1 | 46 | 42 | 42 | 91.30% | 91.30% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CHAFFEE R-II | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CHAFFEE R-II | 1 | 9 | 8 | 8 | 88.89% | 88.89% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CHAFFEE R-II | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CHAFFEE R-II | 1 | 9 | 6 | 6 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G04-Math-N CHAFFEE R-II | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl CHAFFEE R-II | 1 | 31 | 30 | 30 | 96.77% | 96.77% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl CHAFFEE R-II | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl CHAFFEE R-II | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Land Form CHILLICOTHE R-II | 1 | 147 | 105 | 103 | 71.43% | 70.07% |
| 5/19/2014 2:00 | Math-PT-Talent Sh CHILLICOTHE R-II | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh CHILLICOTHE R-II | 1 | 135 | 120 | 120 | 88.89% | 88.89% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No CHILLICOTHE R-II | 1 | 147 | 134 | 123 | 91.16% | 83.67% |
| 5/19/2014 2:00 | SBAC-G06-Math-N CHILLICOTHE R-II | 1 | 30 | 30 | 30 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N CHILLICOTHE R-II | 1 | 28 | 28 | 28 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N CHILLICOTHE R-II | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N CHILLICOTHE R-II | 1 | 37 | 31 | 31 | 83.78% | 83.78% |
| 5/19/2014 2:00 | SBAC-G06-Math-N CHILLICOTHE R-II | 1 | 40 | 34 | 34 | 85.00% | 85.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi CITY GARDEN MOI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | ELA-PT-The Ameri | CITY GARDEN MOI | 1 | 26 | 26 | 26 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-South Po | CITY GARDEN MOI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CITY GARDEN MOI | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CITY GARDEN MOI | 1 | 6 | 6 | 5 | 100.00% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CITY GARDEN MOI | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CITY GARDEN MOI | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | CITY GARDEN MOI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | CITY GARDEN MOI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | CLARK CO. R-I | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | CLARK CO. R-I | 1 | 49 | 46 | 46 | 93.88% | 93.88% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | CLARK CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | CLARK CO. R-I | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | CLARK CO. R-I | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | CLARK CO. R-I | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | CLARK CO. R-I | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | CLARK CO. R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | CLARK CO. R-I | 1 | 49 | 46 | 45 | 93.88% | 91.84% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | CLARK CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | CLARKTON C-4 | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | CLARKTON C-4 | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 | HS-Math-PT-Great | CLEARWATER R-I | 1 | 64 | 61 | 61 | 95.31% | 95.31% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | CLEARWATER R-I | 1 | 64 | 62 | 62 | 96.88% | 96.88% |
| 5/19/2014 2:00 | ELA-PT-Growth an | CLINTON CO. R-III | 1 | 49 | 46 | 45 | 93.88% | 91.84% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CLINTON CO. R-III | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CLINTON CO. R-III | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CLINTON CO. R-III | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | CLINTON CO. R-III | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | COLE CO. R-V | 1 | 42 | 39 | 32 | 92.86% | 76.19% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | COLE CO. R-V | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | COLE CO. R-V | 1 | 43 | 1 | 0 | 2.33% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | COLE CO. R-V | 1 | 42 | 42 | 40 | 100.00% | 95.24% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | COLE CO. R-V | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | COLE CO. R-V | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | COLE CO. R-V | 1 | 43 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | COLUMBIA 93 | 1 | 58 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | COLUMBIA 93 | 1 | 81 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | COLUMBIA 93 | 1 | 404 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | COLUMBIA 93 | 1 | 57 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | COLUMBIA 93 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | COLUMBIA 93 | 1 | 194 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | COLUMBIA 93 | 1 | 31 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | COLUMBIA 93 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | COLUMBIA 93 | 1 | 20 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G03-Math-N COLUMBIA 93 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No COLUMBIA 93 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No COLUMBIA 93 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No COLUMBIA 93 | 1 | 27 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No COLUMBIA 93 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No COLUMBIA 93 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N COLUMBIA 93 | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N COLUMBIA 93 | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N COLUMBIA 93 | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N COLUMBIA 93 | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N COLUMBIA 93 | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N COLUMBIA 93 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N COLUMBIA 93 | 1 | 96 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N COLUMBIA 93 | 1 | 98 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No COLUMBIA 93 | 1 | 404 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great CRANE R-III | 1 | 54 | 52 | 52 | 96.30% | 96.30% |
| 5/19/2014 2:00 | SBAC-HS-Math-No CRANE R-III | 1 | 54 | 52 | 52 | 96.30% | 96.30% |
| 5/19/2014 2:00 | HS-ELA-PT-A New CRAWFORD CO. R- | 1 | 70 | 70 | 70 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New CRAWFORD CO. R- | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great CRAWFORD CO. R- | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non CRAWFORD CO. R- | 1 | 70 | 69 | 69 | 98.57% | 98.57% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non CRAWFORD CO. R- | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No CRAWFORD CO. R- | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New DADEVILLE R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great DADEVILLE R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non DADEVILLE R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No DADEVILLE R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | HS-ELA-PT-A New DALLAS CO. R-I | 1 | 110 | 1 | 0 | 0.91% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New DALLAS CO. R-I | 1 | 12 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great DALLAS CO. R-I | 1 | 122 | 109 | 109 | 89.34% | 89.34% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non DALLAS CO. R-I | 1 | 110 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non DALLAS CO. R-I | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non DALLAS CO. R-I | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No DALLAS CO. R-I | 1 | 122 | 110 | 107 | 90.16% | 87.70% |
| 5/19/2014 2:00 | ELA-PT-Animals W DELTA C-7 | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Technolog DELTA C-7 | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No DELTA C-7 | 1 | 14 | 14 | 9 | 100.00% | 64.29% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No DELTA C-7 | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New DELTA R-V | 1 | 16 | 16 | 16 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non DELTA R-V | 1 | 16 | 16 | 16 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A DESOTO 73 | 1 | 93 | 90 | 90 | 96.77% | 96.77% |
| 5/19/2014 2:00 | SBAC-G06-Math-N DESOTO 73 | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N DESOTO 73 | 1 | 23 | 23 | 22 | 100.00% | 95.65% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G06-Math-N DESOTO 73 | 1 | 24 | 23 | 23 | 95.83% | 95.83% |
| 5/19/2014 2:00 | SBAC-G06-Math-N DESOTO 73 | 1 | 24 | 23 | 23 | 95.83% | 95.83% |
| 5/19/2014 2:00 | HS-ELA-PT-A New DUNKLIN R-V | 1 | 113 | 83 | 83 | 73.45% | 73.45% |
| 5/19/2014 2:00 | Math-PT-Talent Sh DUNKLIN R-V | 1 | 140 | 119 | 119 | 85.00% | 85.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N DUNKLIN R-V | 1 | 36 | 31 | 31 | 86.11% | 86.11% |
| 5/19/2014 2:00 | SBAC-G05-Math-N DUNKLIN R-V | 1 | 34 | 29 | 29 | 85.29% | 85.29% |
| 5/19/2014 2:00 | SBAC-G05-Math-N DUNKLIN R-V | 1 | 36 | 30 | 30 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-Math-N DUNKLIN R-V | 1 | 34 | 29 | 29 | 85.29% | 85.29% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl DUNKLIN R-V | 1 | 78 | 56 | 56 | 71.79% | 71.79% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl DUNKLIN R-V | 1 | 35 | 27 | 26 | 77.14% | 74.29% |
| 5/19/2014 2:00 | HS-ELA-PT-A New EAST CARTER CO. I | 1 | 43 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great EAST CARTER CO. I | 1 | 55 | 43 | 43 | 78.18% | 78.18% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl EAST CARTER CO. I | 1 | 43 | 1 | 1 | 2.33% | 2.33% |
| 5/19/2014 2:00 | SBAC-HS-Math-No EAST CARTER CO. I | 1 | 55 | 43 | 39 | 78.18% | 70.91% |
| 5/19/2014 2:00 | Math-PT-South Po EAST LYNNE 40 | 1 | 16 | 15 | 15 | 93.75% | 93.75% |
| 5/19/2014 2:00 | SBAC-G07-Math-N EAST LYNNE 40 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N EAST LYNNE 40 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N EAST LYNNE 40 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N EAST LYNNE 40 | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great EAST NEWTON CO | 1 | 96 | 94 | 94 | 97.92% | 97.92% |
| 5/19/2014 2:00 | HS-Math-PT-Great EAST NEWTON CO | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No EAST NEWTON CO | 1 | 96 | 96 | 96 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No EAST NEWTON CO | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great EAST PRAIRIE R-II | 1 | 65 | 52 | 52 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No EAST PRAIRIE R-II | 1 | 65 | 52 | 52 | 80.00% | 80.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani EL DORADO SPRIN | 1 | 100 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol EL DORADO SPRIN | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol EL DORADO SPRIN | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol EL DORADO SPRIN | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol EL DORADO SPRIN | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol EL DORADO SPRIN | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po ELDON R-I | 1 | 145 | 134 | 134 | 92.41% | 92.41% |
| 5/19/2014 2:00 | SBAC-G08-Math-N ELDON R-I | 1 | 145 | 135 | 134 | 93.10% | 92.41% |
| 5/19/2014 2:00 | HS-Math-PT-Great EVERTON R-III | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No EVERTON R-III | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New EXETER R-VI | 1 | 31 | 31 | 31 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New EXETER R-VI | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl EXETER R-VI | 1 | 31 | 1 | 1 | 3.23% | 3.23% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl EXETER R-VI | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Camping FAIRFAX R-III | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G07-Math-N FAIRFAX R-III | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N FAIRFAX R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N FAIRFAX R-III | 1 | 4 | 3 | 3 | 75.00% | 75.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G07-Math-N FAIRFAX R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Animals W FARMINGTON R-V | 1 | 93 | 92 | 92 | 98.92% | 98.92% |
| 5/19/2014 2:00 ELA-PT-Marine Ani FARMINGTON R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Commun FARMINGTON R-V | 1 | 295 | 274 | 270 | 92.88% | 91.53% |
| 5/19/2014 2:00 Math-PT-South Po FARMINGTON R-V | 1 | 255 | 247 | 247 | 96.86% | 96.86% |
| 5/19/2014 2:00 Math-PT-Turtle Ha FARMINGTON R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G03-ELA-Nor FARMINGTON R-V | 1 | 31 | 30 | 30 | 96.77% | 96.77% |
| 5/19/2014 2:00 SBAC-G03-ELA-Nor FARMINGTON R-V | 1 | 32 | 32 | 32 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G03-ELA-Nor FARMINGTON R-V | 1 | 30 | 30 | 30 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G05-ELA-Nor FARMINGTON R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G05-Math-N FARMINGTON R-V | 1 | 75 | 72 | 72 | 96.00% | 96.00% |
| 5/19/2014 2:00 SBAC-G05-Math-N FARMINGTON R-V | 1 | 72 | 69 | 69 | 95.83% | 95.83% |
| 5/19/2014 2:00 SBAC-G05-Math-N FARMINGTON R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G05-Math-N FARMINGTON R-V | 1 | 75 | 69 | 69 | 92.00% | 92.00% |
| 5/19/2014 2:00 SBAC-G05-Math-N FARMINGTON R-V | 1 | 73 | 70 | 70 | 95.89% | 95.89% |
| 5/19/2014 2:00 SBAC-G07-Math-N FARMINGTON R-V | 1 | 60 | 60 | 60 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G07-Math-N FARMINGTON R-V | 1 | 63 | 63 | 62 | 100.00% | 98.41% |
| 5/19/2014 2:00 SBAC-G07-Math-N FARMINGTON R-V | 1 | 68 | 64 | 64 | 94.12% | 94.12% |
| 5/19/2014 2:00 SBAC-G07-Math-N FARMINGTON R-V | 1 | 64 | 63 | 63 | 98.44% | 98.44% |
| 5/19/2014 2:00 ELA-PT-Growth an FERGUSON-FLORIS | 1 | 63 | 54 | 54 | 85.71% | 85.71% |
| 5/19/2014 2:00 Math-PT-Cell Phor FERGUSON-FLORIS | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 Math-PT-Cell Phor FERGUSON-FLORIS | 1 | 60 | 54 | 54 | 90.00% | 90.00% |
| 5/19/2014 2:00 Math-PT-Commun FERGUSON-FLORIS | 1 | 54 | 53 | 53 | 98.15% | 98.15% |
| 5/19/2014 2:00 Math-PT-Sandbox- FERGUSON-FLORIS | 1 | 68 | 60 | 60 | 88.24% | 88.24% |
| 5/19/2014 2:00 Math-PT-Turtle Ha FERGUSON-FLORIS | 1 | 63 | 60 | 60 | 95.24% | 95.24% |
| 5/19/2014 2:00 SBAC-G04-Math-N FERGUSON-FLORIS | 1 | 23 | 22 | 21 | 95.65% | 91.30% |
| 5/19/2014 2:00 SBAC-G04-Math-N FERGUSON-FLORIS | 1 | 24 | 24 | 23 | 100.00% | 95.83% |
| 5/19/2014 2:00 SBAC-G04-Math-N FERGUSON-FLORIS | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 SBAC-G04-Math-N FERGUSON-FLORIS | 1 | 24 | 24 | 24 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N FERGUSON-FLORIS | 1 | 23 | 23 | 22 | 100.00% | 95.65% |
| 5/19/2014 2:00 SBAC-G05-ELA-Nor FERGUSON-FLORIS | 1 | 15 | 13 | 13 | 86.67% | 86.67% |
| 5/19/2014 2:00 SBAC-G05-ELA-Nor FERGUSON-FLORIS | 1 | 16 | 15 | 15 | 93.75% | 93.75% |
| 5/19/2014 2:00 SBAC-G05-ELA-Nor FERGUSON-FLORIS | 1 | 16 | 13 | 13 | 81.25% | 81.25% |
| 5/19/2014 2:00 SBAC-G05-ELA-Nor FERGUSON-FLORIS | 1 | 16 | 13 | 13 | 81.25% | 81.25% |
| 5/19/2014 2:00 SBAC-G05-Math-N FERGUSON-FLORIS | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 SBAC-G05-Math-N FERGUSON-FLORIS | 1 | 17 | 13 | 13 | 76.47% | 76.47% |
| 5/19/2014 2:00 SBAC-G05-Math-N FERGUSON-FLORIS | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 SBAC-G05-Math-N FERGUSON-FLORIS | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 SBAC-G06-Math-N FERGUSON-FLORIS | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N FERGUSON-FLORIS | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N FERGUSON-FLORIS | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N FERGUSON-FLORIS | 1 | 15 | 14 | 14 | 93.33% | 93.33% |
| 5/19/2014 2:00 SBAC-G06-Math-N FERGUSON-FLORIS | 1 | 16 | 13 | 13 | 81.25% | 81.25% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great | FESTUS R-VI | 1 | 254 | 59 | 59 | 23.23% | 23.23% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | FESTUS R-VI | 1 | 254 | 59 | 58 | 23.23% | 22.83% |
| 5/19/2014 2:00 | ELA-PT-Animals W | FORT OSAGE R-I | 1 | 71 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | FORT OSAGE R-I | 1 | 79 | 50 | 34 | 63.29% | 43.04% |
| 5/19/2014 2:00 | ELA-PT-Uncommoi | FORT OSAGE R-I | 1 | 191 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommoi | FORT OSAGE R-I | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Jı | FORT OSAGE R-I | 1 | 56 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Jı | FORT OSAGE R-I | 1 | 221 | 54 | 51 | 24.43% | 23.08% |
| 5/19/2014 2:00 | Math-PT-Making S | FORT OSAGE R-I | 1 | 53 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | FORT OSAGE R-I | 1 | 107 | 78 | 76 | 72.90% | 71.03% |
| 5/19/2014 2:00 | Math-PT-School Lil | FORT OSAGE R-I | 1 | 88 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | FORT OSAGE R-I | 1 | 71 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FORT OSAGE R-I | 1 | 50 | 38 | 37 | 76.00% | 74.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FORT OSAGE R-I | 1 | 61 | 48 | 45 | 78.69% | 73.77% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FORT OSAGE R-I | 1 | 58 | 46 | 44 | 79.31% | 75.86% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FORT OSAGE R-I | 1 | 141 | 79 | 74 | 56.03% | 52.48% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 15 | 12 | 11 | 80.00% | 73.33% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 19 | 16 | 13 | 84.21% | 68.42% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 191 | 62 | 53 | 32.46% | 27.75% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 33 | 15 | 13 | 45.45% | 39.39% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 20 | 13 | 11 | 65.00% | 55.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Noi | FORT OSAGE R-I | 1 | 16 | 11 | 10 | 68.75% | 62.50% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FORT OSAGE R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FORT OSAGE R-I | 1 | 56 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FORT OSAGE R-I | 1 | 81 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FORT OSAGE R-I | 1 | 76 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an | FOX C-6 | 1 | 100 | 95 | 95 | 95.00% | 95.00% |
| 5/19/2014 2:00 | ELA-PT-Trees | FOX C-6 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | FOX C-6 | 1 | 85 | 79 | 79 | 92.94% | 92.94% |
| 5/19/2014 2:00 | Math-PT-Commun | FOX C-6 | 1 | 95 | 68 | 66 | 71.58% | 69.47% |
| 5/19/2014 2:00 | Math-PT-South Po | FOX C-6 | 1 | 248 | 202 | 202 | 81.45% | 81.45% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | FOX C-6 | 1 | 29 | 27 | 27 | 93.10% | 93.10% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | FOX C-6 | 1 | 28 | 26 | 25 | 92.86% | 89.29% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | FOX C-6 | 1 | 28 | 27 | 27 | 96.43% | 96.43% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | FOX C-6 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FOX C-6 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FOX C-6 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FOX C-6 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FOX C-6 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FOX C-6 | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Noi | FOX C-6 | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Noi | FOX C-6 | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Noi | FOX C-6 | 1 | 25 | 24 | 24 | 96.00% | 96.00% |

| Date | Assessment | District | | Total | Col4 | Col5 | Pct1 | Pct2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G05-ELA-Non | FOX C-6 | 1 | 25 | 23 | 23 | 92.00% | 92.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FOX C-6 | 1 | 98 | 75 | 75 | 76.53% | 76.53% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FOX C-6 | 1 | 100 | 84 | 84 | 84.00% | 84.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FOX C-6 | 1 | 50 | 43 | 43 | 86.00% | 86.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | FRANCIS HOWELL | 1 | 276 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | FRANCIS HOWELL | 1 | 105 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FRANCIS HOWELL | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FRANCIS HOWELL | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FRANCIS HOWELL | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FRANCIS HOWELL | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | FRANCIS HOWELL | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FRANCIS HOWELL | 1 | 110 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FRANCIS HOWELL | 1 | 110 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | FRANCIS HOWELL | 1 | 56 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | FRONTIER SCHOOL | 1 | 34 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | FRONTIER SCHOOL | 1 | 34 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | FULTON 58 | 1 | 168 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | FULTON 58 | 1 | 147 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Commun | FULTON 58 | 1 | 71 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | FULTON 58 | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FULTON 58 | 1 | 14 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FULTON 58 | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | FULTON 58 | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | FULTON 58 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | FULTON 58 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | FULTON 58 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | FULTON 58 | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Non | FULTON 58 | 1 | 42 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Non | FULTON 58 | 1 | 42 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Non | FULTON 58 | 1 | 42 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Non | FULTON 58 | 1 | 42 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | FULTON 58 | 1 | 147 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | GALLATIN R-V | 1 | 48 | 48 | 48 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | GALLATIN R-V | 1 | 48 | 48 | 48 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | GASCONADE CO. R | 1 | 49 | 46 | 46 | 93.88% | 93.88% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 12 | 9 | 9 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | GASCONADE CO. R | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | GASCONADE CO. R | 1 | 103 | 102 | 102 | 99.03% | 99.03% |
| 5/19/2014 2:00 | HS-Math-PT-Great | GASCONADE CO. R | 1 | 128 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Cell Phon | GASCONADE CO. R | 1 | 133 | 132 | 132 | 99.25% | 99.25% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | Math-PT-Talent Sh | GASCONADE CO. R | 1 | 38 | 36 | 36 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GASCONADE CO. R | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GASCONADE CO. R | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GASCONADE CO. R | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GASCONADE CO. R | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GASCONADE CO. R | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 33 | 33 | 33 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 32 | 32 | 32 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 34 | 33 | 33 | 97.06% | 97.06% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GASCONADE CO. R | 1 | 34 | 34 | 34 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | GASCONADE CO. R | 1 | 128 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | GATEWAY SCIENC | 1 | 44 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | GATEWAY SCIENC | 1 | 44 | 40 | 25 | 90.91% | 56.82% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | GILMAN CITY R-IV | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GILMAN CITY R-IV | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GILMAN CITY R-IV | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | GILMAN CITY R-IV | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | GOLDEN CITY R-III | 1 | 18 | 15 | 15 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | GOLDEN CITY R-III | 1 | 18 | 15 | 15 | 83.33% | 83.33% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | GRAIN VALLEY R-V | 1 | 89 | 89 | 89 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GRAIN VALLEY R-V | 1 | 22 | 22 | 20 | 100.00% | 90.91% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GRAIN VALLEY R-V | 1 | 22 | 22 | 21 | 100.00% | 95.45% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GRAIN VALLEY R-V | 1 | 22 | 22 | 21 | 100.00% | 95.45% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GRAIN VALLEY R-V | 1 | 23 | 23 | 21 | 100.00% | 91.30% |
| 5/19/2014 2:00 | ELA-PT-Importanc | GRAND CENTER AF | 1 | 86 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | GRAND CENTER AF | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | GRAND CENTER AF | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | GRAND CENTER AF | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | GRAND CENTER AF | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | GRANDVIEW C-4 | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | GRANDVIEW C-4 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | GRANDVIEW C-4 | 1 | 76 | 70 | 70 | 92.11% | 92.11% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | GRANDVIEW C-4 | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | GRANDVIEW C-4 | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | GRANDVIEW C-4 | 1 | 25 | 21 | 21 | 84.00% | 84.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | GRANDVIEW C-4 | 1 | 26 | 24 | 24 | 92.31% | 92.31% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | GRANDVIEW C-4 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | GRANDVIEW R-II | 1 | 1 | 1 | 1 | 100.00% | 100.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | ELA-PT-Uncommo | GRANDVIEW R-II | 1 | 57 | 53 | 53 | 92.98% | 92.98% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 12 | 12 | 11 | 100.00% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 11 | 11 | 11 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 12 | 12 | 11 | 100.00% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRANDVIEW R-II | 1 | 11 | 10 | 9 | 90.91% | 81.82% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | GREEN FOREST R-I | 1 | 22 | 20 | 20 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GREEN FOREST R-I | 1 | 6 | 5 | 5 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GREEN FOREST R-I | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GREEN FOREST R-I | 1 | 6 | 5 | 5 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | GREEN FOREST R-I | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | GREENFIELD R-IV | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | GREENFIELD R-IV | 1 | 14 | 14 | 12 | 100.00% | 85.71% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | GREENFIELD R-IV | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | Math-PT-Making S | GREENVILLE R-II | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | GREENVILLE R-II | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | ELA-PT-Marine An | GRUNDY CO. R-V | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Camping | GRUNDY CO. R-V | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRUNDY CO. R-V | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRUNDY CO. R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRUNDY CO. R-V | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRUNDY CO. R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | GRUNDY CO. R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | GRUNDY CO. R-V | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | GRUNDY CO. R-V | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | GRUNDY CO. R-V | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | GRUNDY CO. R-V | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | HALE R-I | 1 | 14 | 9 | 7 | 64.29% | 50.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | HALE R-I | 1 | 3 | 1 | 0 | 33.33% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | HALE R-I | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | HALE R-I | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | HALE R-I | 1 | 3 | 1 | 0 | 33.33% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | HALE R-I | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | HALLSVILLE R-IV | 1 | 107 | 1 | 1 | 0.93% | 0.93% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | HALLSVILLE R-IV | 1 | 107 | 96 | 61 | 89.72% | 57.01% |
| 5/19/2014 2:00 | Math-PT-Cell Phor | HAMILTON R-II | 1 | 41 | 41 | 41 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | HAMILTON R-II | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | HAMILTON R-II | 1 | 20 | 20 | 20 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | HANNIBAL 60 | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | HANNIBAL 60 | 1 | 253 | 84 | 62 | 33.20% | 24.51% |
| 5/19/2014 2:00 | Math-PT-Order Fo | HANNIBAL 60 | 1 | 74 | 72 | 70 | 97.30% | 94.59% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HANNIBAL 60 | 1 | 15 | 14 | 14 | 93.33% | 93.33% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G04-Math-N | HANNIBAL 60 | 1 | 15 | 14 | 14 | 93.33% | 93.33% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HANNIBAL 60 | 1 | 15 | 15 | 14 | 100.00% | 93.33% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HANNIBAL 60 | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HANNIBAL 60 | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | HANNIBAL 60 | 1 | 64 | 21 | 0 | 32.81% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | HANNIBAL 60 | 1 | 63 | 28 | 5 | 44.44% | 7.94% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | HANNIBAL 60 | 1 | 3 | 2 | 1 | 66.67% | 33.33% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | HANNIBAL 60 | 1 | 63 | 25 | 1 | 39.68% | 1.59% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | HANNIBAL 60 | 1 | 63 | 14 | 2 | 22.22% | 3.17% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | HARDIN-CENTRAL | 1 | 16 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | HARDIN-CENTRAL | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | HARDIN-CENTRAL | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | HARRISBURG R-VII | 1 | 41 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | HARRISBURG R-VII | 1 | 41 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | HAYTI R-II | 1 | 52 | 40 | 40 | 76.92% | 76.92% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | HAYTI R-II | 1 | 52 | 40 | 40 | 76.92% | 76.92% |
| 5/19/2014 2:00 | ELA-PT-Animals W | HAZELWOOD | 1 | 156 | 151 | 148 | 96.79% | 94.87% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | HAZELWOOD | 1 | 204 | 183 | 138 | 89.71% | 67.65% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | HAZELWOOD | 1 | 22 | 21 | 17 | 95.45% | 77.27% |
| 5/19/2014 2:00 | ELA-PT-Growth an | HAZELWOOD | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | HAZELWOOD | 1 | 75 | 64 | 60 | 85.33% | 80.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | HAZELWOOD | 1 | 60 | 49 | 47 | 81.67% | 78.33% |
| 5/19/2014 2:00 | ELA-PT-Uncommor | HAZELWOOD | 1 | 266 | 192 | 180 | 72.18% | 67.67% |
| 5/19/2014 2:00 | ELA-PT-Uncommor | HAZELWOOD | 1 | 38 | 33 | 29 | 86.84% | 76.32% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | HAZELWOOD | 1 | 1020 | 543 | 532 | 53.24% | 52.16% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | HAZELWOOD | 1 | 37 | 34 | 34 | 91.89% | 91.89% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | HAZELWOOD | 1 | 96 | 91 | 90 | 94.79% | 93.75% |
| 5/19/2014 2:00 | Math-PT-Donuts | HAZELWOOD | 1 | 290 | 257 | 244 | 88.62% | 84.14% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | HAZELWOOD | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | Math-PT-Making S | HAZELWOOD | 1 | 101 | 46 | 46 | 45.54% | 45.54% |
| 5/19/2014 2:00 | Math-PT-Making S | HAZELWOOD | 1 | 39 | 14 | 14 | 35.90% | 35.90% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | HAZELWOOD | 1 | 387 | 342 | 334 | 88.37% | 86.30% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | HAZELWOOD | 1 | 113 | 96 | 94 | 84.96% | 83.19% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor | HAZELWOOD | 1 | 156 | 151 | 149 | 96.79% | 95.51% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | HAZELWOOD | 1 | 15 | 8 | 8 | 53.33% | 53.33% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | HAZELWOOD | 1 | 24 | 7 | 7 | 29.17% | 29.17% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | HAZELWOOD | 1 | 101 | 47 | 47 | 46.53% | 46.53% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | HAZELWOOD | 1 | 9 | 9 | 8 | 100.00% | 88.89% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | HAZELWOOD | 1 | 266 | 253 | 220 | 95.11% | 82.71% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | HAZELWOOD | 1 | 16 | 14 | 14 | 87.50% | 87.50% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | HAZELWOOD | 1 | 13 | 12 | 10 | 92.31% | 76.92% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HAZELWOOD | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | HAZELWOOD | 1 | 37 | 34 | 33 | 91.89% | 89.19% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G04-Math-N HAZELWOOD | 1 | 48 | 48 | 46 | 100.00% | 95.83% |
| 5/19/2014 2:00 SBAC-G04-Math-N HAZELWOOD | 1 | 47 | 46 | 43 | 97.87% | 91.49% |
| 5/19/2014 2:00 SBAC-G05-ELA-No HAZELWOOD | 1 | 15 | 14 | 13 | 93.33% | 86.67% |
| 5/19/2014 2:00 SBAC-G05-ELA-No HAZELWOOD | 1 | 23 | 20 | 17 | 86.96% | 73.91% |
| 5/19/2014 2:00 SBAC-G05-ELA-No HAZELWOOD | 1 | 75 | 73 | 67 | 97.33% | 89.33% |
| 5/19/2014 2:00 SBAC-G05-ELA-No HAZELWOOD | 1 | 11 | 11 | 10 | 100.00% | 90.91% |
| 5/19/2014 2:00 SBAC-G05-ELA-No HAZELWOOD | 1 | 11 | 10 | 8 | 90.91% | 72.73% |
| 5/19/2014 2:00 SBAC-G06-ELA-No HAZELWOOD | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N HAZELWOOD | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N HAZELWOOD | 1 | 26 | 26 | 24 | 100.00% | 92.31% |
| 5/19/2014 2:00 SBAC-G06-Math-N HAZELWOOD | 1 | 387 | 355 | 334 | 91.73% | 86.30% |
| 5/19/2014 2:00 SBAC-G06-Math-N HAZELWOOD | 1 | 38 | 33 | 28 | 86.84% | 73.68% |
| 5/19/2014 2:00 SBAC-G06-Math-N HAZELWOOD | 1 | 42 | 34 | 33 | 80.95% | 78.57% |
| 5/19/2014 2:00 SBAC-G07-Math-N HAZELWOOD | 1 | 290 | 258 | 221 | 88.97% | 76.21% |
| 5/19/2014 2:00 SBAC-G07-Math-N HAZELWOOD | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G07-Math-N HAZELWOOD | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-ELA-No HAZELWOOD | 1 | 1 | 1 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G08-ELA-No HAZELWOOD | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-ELA-No HAZELWOOD | 1 | 204 | 193 | 157 | 94.61% | 76.96% |
| 5/19/2014 2:00 SBAC-G08-ELA-No HAZELWOOD | 1 | 7 | 6 | 4 | 85.71% | 57.14% |
| 5/19/2014 2:00 SBAC-G08-ELA-No HAZELWOOD | 1 | 11 | 11 | 7 | 100.00% | 63.64% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HAZELWOOD | 1 | 570 | 300 | 273 | 52.63% | 47.89% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HAZELWOOD | 1 | 450 | 280 | 275 | 62.22% | 61.11% |
| 5/19/2014 2:00 HS-ELA-PT-A New HENRY CO. R-I | 1 | 59 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-Math-PT-Great HENRY CO. R-I | 1 | 59 | 59 | 59 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HENRY CO. R-I | 1 | 59 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-Math-No HENRY CO. R-I | 1 | 59 | 59 | 58 | 100.00% | 98.31% |
| 5/19/2014 2:00 Math-PT-South Po HIGBEE R-VIII | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N HIGBEE R-VIII | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N HIGBEE R-VIII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N HIGBEE R-VIII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N HIGBEE R-VIII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 HS-ELA-PT-A New HILLSBORO R-III | 1 | 54 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HILLSBORO R-III | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HILLSBORO R-III | 1 | 19 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-ELA-PT-A New HOLCOMB R-III | 1 | 39 | 32 | 32 | 82.05% | 82.05% |
| 5/19/2014 2:00 HS-ELA-PT-A New HOLCOMB R-III | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-Math-PT-Great HOLCOMB R-III | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HOLCOMB R-III | 1 | 39 | 39 | 39 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HOLCOMB R-III | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Non HOLCOMB R-III | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-Math-No HOLCOMB R-III | 1 | 39 | 1 | 0 | 2.56% | 0.00% |

| Date | Assessment | District | | Count1 | Count2 | Count3 | Pct1 | Pct2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-ELA-PT-A New | HOLDEN R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | HOLDEN R-III | 1 | 111 | 105 | 97 | 94.59% | 87.39% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 23 | 19 | 18 | 82.61% | 78.26% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 21 | 21 | 15 | 100.00% | 71.43% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 23 | 23 | 21 | 100.00% | 91.30% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 22 | 22 | 19 | 100.00% | 86.36% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | HOLDEN R-III | 1 | 22 | 22 | 16 | 100.00% | 72.73% |
| 5/19/2014 2:00 | HS-Math-PT-Great | HUMANSVILLE R-IV | 1 | 25 | 25 | 25 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | HUMANSVILLE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | HUMANSVILLE R-IV | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | HUMANSVILLE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | HUMANSVILLE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | HUMANSVILLE R-IV | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-South Po | IBERIA R-V | 1 | 39 | 37 | 37 | 94.87% | 94.87% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | IBERIA R-V | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | IBERIA R-V | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | IBERIA R-V | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | IBERIA R-V | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an | INDEPENDENCE 3C | 1 | 77 | 74 | 69 | 96.10% | 89.61% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | INDEPENDENCE 3C | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | INDEPENDENCE 3C | 1 | 100 | 98 | 96 | 98.00% | 96.00% |
| 5/19/2014 2:00 | ELA-PT-The Ameri | INDEPENDENCE 3C | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-The Ameri | INDEPENDENCE 3C | 1 | 312 | 140 | 139 | 44.87% | 44.55% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | INDEPENDENCE 3C | 1 | 79 | 71 | 69 | 89.87% | 87.34% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | INDEPENDENCE 3C | 1 | 73 | 67 | 67 | 91.78% | 91.78% |
| 5/19/2014 2:00 | Math-PT-Order Fo | INDEPENDENCE 3C | 1 | 42 | 41 | 41 | 97.62% | 97.62% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | INDEPENDENCE 3C | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | INDEPENDENCE 3C | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | INDEPENDENCE 3C | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | INDEPENDENCE 3C | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | INDEPENDENCE 3C | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | INDEPENDENCE 3C | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | INDEPENDENCE 3C | 1 | 22 | 22 | 21 | 100.00% | 95.45% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | INDEPENDENCE 3C | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | INDEPENDENCE 3C | 1 | 24 | 23 | 22 | 95.83% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | INDEPENDENCE 3C | 1 | 15 | 12 | 12 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | INDEPENDENCE 3C | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | INDEPENDENCE 3C | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | INDEPENDENCE 3C | 1 | 15 | 14 | 14 | 93.33% | 93.33% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | INDEPENDENCE 3C | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | INDEPENDENCE 3C | 1 | 37 | 37 | 37 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | INDEPENDENCE 3C | 1 | 36 | 34 | 33 | 94.44% | 91.67% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G05-ELA-No | INDEPENDENCE 3( | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | INDEPENDENCE 3( | 1 | 37 | 36 | 35 | 97.30% | 94.59% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | INDEPENDENCE 3( | 1 | 37 | 37 | 37 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | INDEPENDENCE 3( | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | INDEPENDENCE 3( | 1 | 78 | 36 | 33 | 46.15% | 42.31% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | INDEPENDENCE 3( | 1 | 78 | 34 | 33 | 43.59% | 42.31% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | INDEPENDENCE 3( | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | INDEPENDENCE 3( | 1 | 78 | 38 | 38 | 48.72% | 48.72% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | INDEPENDENCE 3( | 1 | 78 | 36 | 35 | 46.15% | 44.87% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | IRON CO. C-4 | 1 | 28 | 25 | 25 | 89.29% | 89.29% |
| 5/19/2014 2:00 | HS-Math-PT-Great | IRON CO. C-4 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | IRON CO. C-4 | 1 | 28 | 28 | 28 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | IRON CO. C-4 | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an | JACKSON R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Inventions | JACKSON R-II | 1 | 352 | 338 | 338 | 96.02% | 96.02% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | JACKSON R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | JACKSON R-II | 1 | 355 | 342 | 341 | 96.34% | 96.06% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | JACKSON R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | JAMESTOWN C-1 | 1 | 21 | 17 | 15 | 80.95% | 71.43% |
| 5/19/2014 2:00 | Math-PT-Cell Phor | JAMESTOWN C-1 | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | JAMESTOWN C-1 | 1 | 4 | 4 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | JAMESTOWN C-1 | 1 | 4 | 4 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | JAMESTOWN C-1 | 1 | 4 | 4 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | JAMESTOWN C-1 | 1 | 5 | 4 | 0 | 80.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | JAMESTOWN C-1 | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | JAMESTOWN C-1 | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | JAMESTOWN C-1 | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | JAMESTOWN C-1 | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | JAMESTOWN C-1 | 1 | 4 | 4 | 3 | 100.00% | 75.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | JEFFERSON CITY | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | JEFFERSON CITY | 1 | 72 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Technology | JEFFERSON CITY | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Technology | JEFFERSON CITY | 1 | 301 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | JEFFERSON CITY | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | JEFFERSON CITY | 1 | 76 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | JEFFERSON CITY | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | JEFFERSON CITY | 1 | 84 | 39 | 36 | 46.43% | 42.86% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 14 | 14 | 13 | 100.00% | 92.86% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 14 | 11 | 8 | 78.57% | 57.14% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 4 | 4 | 3 | 100.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 15 | 13 | 11 | 86.67% | 73.33% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 14 | 13 | 8 | 92.86% | 57.14% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | JEFFERSON CITY | 1 | 15 | 14 | 11 | 93.33% | 73.33% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 31 | 23 | 20 | 74.19% | 64.52% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 33 | 24 | 24 | 72.73% | 72.73% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 5 | 3 | 3 | 60.00% | 60.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 32 | 24 | 24 | 75.00% | 75.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 32 | 20 | 20 | 62.50% | 62.50% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CITY | 1 | 32 | 25 | 23 | 78.13% | 71.88% |
| 5/19/2014 2:00 SBAC-G07-ELA-No JEFFERSON CITY | 1 | 76 | 72 | 69 | 94.74% | 90.79% |
| 5/19/2014 2:00 SBAC-G07-ELA-No JEFFERSON CITY | 1 | 75 | 66 | 66 | 88.00% | 88.00% |
| 5/19/2014 2:00 SBAC-G07-ELA-No JEFFERSON CITY | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G07-ELA-No JEFFERSON CITY | 1 | 75 | 70 | 69 | 93.33% | 92.00% |
| 5/19/2014 2:00 SBAC-G07-ELA-No JEFFERSON CITY | 1 | 75 | 70 | 70 | 93.33% | 93.33% |
| 5/19/2014 2:00 Math-PT-Animal Ju JEFFERSON CO. R-\ | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Animal Ju JEFFERSON CO. R-\ | 1 | 87 | 75 | 75 | 86.21% | 86.21% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 17 | 14 | 12 | 82.35% | 70.59% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 18 | 16 | 16 | 88.89% | 88.89% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 SBAC-G04-Math-N JEFFERSON CO. R-\ | 1 | 18 | 17 | 14 | 94.44% | 77.78% |
| 5/19/2014 2:00 HS-Math-PT-Great JENNINGS | 1 | 87 | 69 | 69 | 79.31% | 79.31% |
| 5/19/2014 2:00 SBAC-HS-Math-No JENNINGS | 1 | 87 | 72 | 72 | 82.76% | 82.76% |
| 5/19/2014 2:00 ELA-PT-Growth an JOPLIN SCHOOLS | 1 | 400 | 311 | 309 | 77.75% | 77.25% |
| 5/19/2014 2:00 ELA-PT-Marine Ani JOPLIN SCHOOLS | 1 | 114 | 61 | 61 | 53.51% | 53.51% |
| 5/19/2014 2:00 ELA-PT-Trees-A    JOPLIN SCHOOLS | 1 | 41 | 37 | 37 | 90.24% | 90.24% |
| 5/19/2014 2:00 Math-PT-Animal Ju JOPLIN SCHOOLS | 1 | 57 | 52 | 52 | 91.23% | 91.23% |
| 5/19/2014 2:00 Math-PT-Commun JOPLIN SCHOOLS | 1 | 83 | 72 | 72 | 86.75% | 86.75% |
| 5/19/2014 2:00 Math-PT-Order Fo JOPLIN SCHOOLS | 1 | 53 | 49 | 49 | 92.45% | 92.45% |
| 5/19/2014 2:00 SBAC-G03-ELA-No JOPLIN SCHOOLS | 1 | 13 | 11 | 11 | 84.62% | 84.62% |
| 5/19/2014 2:00 SBAC-G03-ELA-No JOPLIN SCHOOLS | 1 | 14 | 12 | 12 | 85.71% | 85.71% |
| 5/19/2014 2:00 SBAC-G03-ELA-No JOPLIN SCHOOLS | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G03-Math-N JOPLIN SCHOOLS | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G03-Math-N JOPLIN SCHOOLS | 1 | 18 | 16 | 16 | 88.89% | 88.89% |
| 5/19/2014 2:00 SBAC-G03-Math-N JOPLIN SCHOOLS | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N JOPLIN SCHOOLS | 1 | 28 | 26 | 26 | 92.86% | 92.86% |
| 5/19/2014 2:00 SBAC-G04-Math-N JOPLIN SCHOOLS | 1 | 28 | 25 | 25 | 89.29% | 89.29% |
| 5/19/2014 2:00 SBAC-G04-Math-N JOPLIN SCHOOLS | 1 | 28 | 25 | 25 | 89.29% | 89.29% |
| 5/19/2014 2:00 SBAC-G04-Math-N JOPLIN SCHOOLS | 1 | 28 | 24 | 24 | 85.71% | 85.71% |
| 5/19/2014 2:00 SBAC-G04-Math-N JOPLIN SCHOOLS | 1 | 28 | 25 | 25 | 89.29% | 89.29% |
| 5/19/2014 2:00 SBAC-G05-ELA-No JOPLIN SCHOOLS | 1 | 114 | 61 | 60 | 53.51% | 52.63% |
| 5/19/2014 2:00 SBAC-G06-ELA-No JOPLIN SCHOOLS | 1 | 100 | 80 | 76 | 80.00% | 76.00% |
| 5/19/2014 2:00 SBAC-G06-ELA-No JOPLIN SCHOOLS | 1 | 100 | 75 | 73 | 75.00% | 73.00% |
| 5/19/2014 2:00 SBAC-G06-ELA-No JOPLIN SCHOOLS | 1 | 100 | 83 | 80 | 83.00% | 80.00% |
| 5/19/2014 2:00 SBAC-G06-ELA-No JOPLIN SCHOOLS | 1 | 100 | 84 | 82 | 84.00% | 82.00% |

| Date | Test | District | | Total | Col1 | Col2 | Pct1 | Pct2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great | KANSAS CITY 33 | 1 | 302 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Commun | KANSAS CITY 33 | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | KANSAS CITY 33 | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | KANSAS CITY 33 | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | KANSAS CITY 33 | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | KANSAS CITY 33 | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | KANSAS CITY 33 | 1 | 302 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Trees | KEARNEY R-I | 1 | 53 | 38 | 28 | 71.70% | 52.83% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | KEARNEY R-I | 1 | 53 | 52 | 47 | 98.11% | 88.68% |
| 5/19/2014 2:00 | Math-PT-South Po | KELSO C-7 | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KELSO C-7 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KELSO C-7 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KELSO C-7 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Heatwaves | KENNETT 39 | 1 | 177 | 161 | 161 | 90.96% | 90.96% |
| 5/19/2014 2:00 | HS-Math-PT-Great | KENNETT 39 | 1 | 128 | 102 | 102 | 79.69% | 79.69% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | KENNETT 39 | 1 | 177 | 166 | 165 | 93.79% | 93.22% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | KENNETT 39 | 1 | 128 | 108 | 108 | 84.38% | 84.38% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | KEYTESVILLE R-III | 1 | 16 | 16 | 16 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | KEYTESVILLE R-III | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | KEYTESVILLE R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | KEYTESVILLE R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | KEYTESVILLE R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | KEYTESVILLE R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | KINGSVILLE R-I | 1 | 20 | 20 | 20 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | KINGSVILLE R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | KINGSVILLE R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | KINGSVILLE R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | KINGSVILLE R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-South Po | KIPP ST LOUIS | 1 | 65 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIPP ST LOUIS | 1 | 65 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Trees | KIRKSVILLE R-III | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | KIRKSVILLE R-III | 1 | 172 | 152 | 152 | 88.37% | 88.37% |
| 5/19/2014 2:00 | Math-PT-Baseball | KIRKSVILLE R-III | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | Math-PT-Baseball- | KIRKSVILLE R-III | 1 | 192 | 152 | 152 | 79.17% | 79.17% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 35 | 28 | 28 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 34 | 29 | 29 | 85.29% | 85.29% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 35 | 32 | 32 | 91.43% | 91.43% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 34 | 31 | 31 | 91.18% | 91.18% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | KIRKSVILLE R-III | 1 | 34 | 32 | 32 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 3 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 75 | 64 | 64 | 85.33% | 85.33% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 73 | 62 | 62 | 84.93% | 84.93% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | KIRKSVILLE R-III | 1 | 39 | 28 | 28 | 71.79% | 71.79% |
| 5/19/2014 2:00 | ELA-PT-Animals W | KNOB NOSTER R-V | 1 | 58 | 55 | 55 | 94.83% | 94.83% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor | KNOB NOSTER R-V | 1 | 58 | 54 | 53 | 93.10% | 91.38% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | KNOX CO. R-I | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | KNOX CO. R-I | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LA MONTE R-IV | 1 | 22 | 21 | 20 | 95.45% | 90.91% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LA MONTE R-IV | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LA MONTE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LA MONTE R-IV | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LA MONTE R-IV | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LA MONTE R-IV | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | LA PLATA R-II | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | LA PLATA R-II | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Fitness C | LACLEDE CO. C-5 | 1 | 38 | 38 | 38 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LACLEDE CO. C-5 | 1 | 38 | 38 | 38 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LACLEDE CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LACLEDE CO. R-I | 1 | 59 | 56 | 56 | 94.92% | 94.92% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LACLEDE CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LACLEDE CO. R-I | 1 | 32 | 31 | 31 | 96.88% | 96.88% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LACLEDE CO. R-I | 1 | 27 | 26 | 26 | 96.30% | 96.30% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LAFAYETTE CO. C-1 | 1 | 154 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Science F | LAFAYETTE CO. C-1 | 1 | 77 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LAFAYETTE CO. C-1 | 1 | 77 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAFAYETTE CO. C-1 | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAFAYETTE CO. C-1 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAFAYETTE CO. C-1 | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAFAYETTE CO. C-1 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAFAYETTE CO. C-1 | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LAFAYETTE CO. C-1 | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LAFAYETTE CO. C-1 | 1 | 31 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LAMAR R-I | 1 | 199 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAMAR R-I | 1 | 22 | 21 | 19 | 95.45% | 86.36% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAMAR R-I | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAMAR R-I | 1 | 21 | 21 | 20 | 100.00% | 95.24% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAMAR R-I | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LAMAR R-I | 1 | 21 | 21 | 19 | 100.00% | 90.48% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LAMAR R-I | 1 | 50 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LAMAR R-I | 1 | 41 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LAQUEY R-V | 1 | 48 | 34 | 30 | 70.83% | 62.50% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | LAQUEY R-V | 1 | 48 | 34 | 34 | 70.83% | 70.83% |

| Date | Assessment | District | | Count 1 | Count 2 | Count 3 | Percent 1 | Percent 2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | ELA-PT-Growth an | LATHROP R-II | 1 | 92 | 88 | 88 | 95.65% | 95.65% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LATHROP R-II | 1 | 60 | 50 | 50 | 83.33% | 83.33% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LATHROP R-II | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor | LATHROP R-II | 1 | 24 | 23 | 22 | 95.83% | 91.67% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor | LATHROP R-II | 1 | 23 | 23 | 22 | 100.00% | 95.65% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor | LATHROP R-II | 1 | 23 | 23 | 23 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor | LATHROP R-II | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | LATHROP R-II | 1 | 60 | 58 | 58 | 96.67% | 96.67% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | LATHROP R-II | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | LATHROP R-II | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | LEWIS CO. C-1 | 1 | 79 | 71 | 70 | 89.87% | 88.61% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | LEWIS CO. C-1 | 1 | 68 | 62 | 47 | 91.18% | 69.12% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | LEWIS CO. C-1 | 1 | 68 | 63 | 61 | 92.65% | 89.71% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LEWIS CO. C-1 | 1 | 79 | 78 | 77 | 98.73% | 97.47% |
| 5/19/2014 2:00 | Math-PT-Making S | LEXINGTON R-V | 1 | 72 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LEXINGTON R-V | 1 | 24 | 19 | 18 | 79.17% | 75.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LEXINGTON R-V | 1 | 25 | 23 | 22 | 92.00% | 88.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LEXINGTON R-V | 1 | 23 | 23 | 22 | 100.00% | 95.65% |
| 5/19/2014 2:00 | HS-Math-PT-Great | LIBERAL R-II | 1 | 35 | 18 | 18 | 51.43% | 51.43% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | LIBERAL R-II | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | LIBERTY 53 | 1 | 96 | 85 | 84 | 88.54% | 87.50% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LIBERTY 53 | 1 | 31 | 30 | 30 | 96.77% | 96.77% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LIBERTY 53 | 1 | 33 | 29 | 29 | 87.88% | 87.88% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | LIBERTY 53 | 1 | 32 | 30 | 26 | 93.75% | 81.25% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LICKING R-VIII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LICKING R-VIII | 1 | 61 | 57 | 57 | 93.44% | 93.44% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | LICKING R-VIII | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | HS-Math-PT-Great | LIFT FOR LIFE ACAI | 1 | 78 | 71 | 71 | 91.03% | 91.03% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | LIFT FOR LIFE ACAI | 1 | 78 | 69 | 68 | 88.46% | 87.18% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | LINCOLN R-II | 1 | 31 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | LINCOLN R-II | 1 | 31 | 31 | 31 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI | LINCOLN R-II | 1 | 31 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | LINCOLN R-II | 1 | 31 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Camping | LINDBERGH SCHOC | 1 | 200 | 181 | 178 | 90.50% | 89.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | LINDBERGH SCHOC | 1 | 80 | 72 | 72 | 90.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | LINDBERGH SCHOC | 1 | 80 | 72 | 72 | 90.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | LINDBERGH SCHOC | 1 | 40 | 39 | 38 | 97.50% | 95.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | LINN CO. R-I | 1 | 18 | 15 | 15 | 83.33% | 83.33% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-HS-Math-No | LINN CO. R-I | 1 | 18 | 15 | 15 | 83.33% | 83.33% |
| 5/19/2014 2:00 | ELA-PT-Renewable | LOCKWOOD R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | LOCKWOOD R-I | 1 | 34 | 31 | 30 | 91.18% | 88.24% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | LOCKWOOD R-I | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | LONE JACK C-6 | 1 | 44 | 43 | 40 | 97.73% | 90.91% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | LONE JACK C-6 | 1 | 5 | 5 | 4 | 100.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | LONE JACK C-6 | 1 | 44 | 42 | 41 | 95.45% | 93.18% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | LONE JACK C-6 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | LONE JACK C-6 | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Trees | LOUISIANA R-II | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | LOUISIANA R-II | 1 | 66 | 56 | 56 | 84.85% | 84.85% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | LOUISIANA R-II | 1 | 21 | 15 | 15 | 71.43% | 71.43% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | LOUISIANA R-II | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | LOUISIANA R-II | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | LOUISIANA R-II | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | MACON CO. R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | MACON CO. R-I | 1 | 101 | 93 | 91 | 92.08% | 90.10% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 21 | 19 | 19 | 90.48% | 90.48% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 20 | 19 | 17 | 95.00% | 85.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 20 | 20 | 18 | 100.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 20 | 20 | 19 | 100.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MACON CO. R-I | 1 | 20 | 19 | 19 | 95.00% | 95.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MACON CO. R-IV | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MACON CO. R-IV | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Importanc | MALDEN R-I | 1 | 72 | 58 | 55 | 80.56% | 76.39% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | MALDEN R-I | 1 | 72 | 58 | 58 | 80.56% | 80.56% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MAPLEWOOD-RICH | 1 | 91 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | MAPLEWOOD-RICH | 1 | 86 | 69 | 69 | 80.23% | 80.23% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | MAPLEWOOD-RICH | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Making S | MAPLEWOOD-RICH | 1 | 68 | 57 | 57 | 83.82% | 83.82% |
| 5/19/2014 2:00 | Math-PT-Making S | MAPLEWOOD-RICH | 1 | 41 | 34 | 34 | 82.93% | 82.93% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MAPLEWOOD-RICH | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MAPLEWOOD-RICH | 1 | 27 | 23 | 23 | 85.19% | 85.19% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MAPLEWOOD-RICH | 1 | 68 | 53 | 52 | 77.94% | 76.47% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MAPLEWOOD-RICH | 1 | 86 | 73 | 72 | 84.88% | 83.72% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MAPLEWOOD-RICH | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MAPLEWOOD-RICH | 1 | 91 | 0 | 0 | 0.00% | 0.00% |

| Date | Assessment | District | | Total | Started | Completed | % Started | % Completed |
|------|-----------|----------|---|-------|---------|-----------|-----------|-------------|
| 5/19/2014 2:00 | Math-PT-Donuts | MARCELINE R-V | 1 | 45 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | MARCELINE R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARCELINE R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARCELINE R-V | 1 | 45 | 43 | 43 | 95.56% | 95.56% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | MARIES CO. R-I | 1 | 23 | 23 | 23 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MARIES CO. R-I | 1 | 47 | 43 | 43 | 91.49% | 91.49% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | MARIES CO. R-I | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MARIES CO. R-I | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | MARIES CO. R-I | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MARIES CO. R-I | 1 | 47 | 46 | 46 | 97.87% | 97.87% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | MARIES CO. R-II | 1 | 53 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | MARIES CO. R-II | 1 | 63 | 50 | 50 | 79.37% | 79.37% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | MARIES CO. R-II | 1 | 53 | 1 | 0 | 1.89% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | MARIES CO. R-II | 1 | 63 | 53 | 53 | 84.13% | 84.13% |
| 5/19/2014 2:00 | ELA-PT-Renewable | MARION CO. R-II | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | MARION CO. R-II | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | MARION CO. R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | MARION CO. R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | MARION CO. R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Technology | MARIONVILLE R-IX | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MARIONVILLE R-IX | 1 | 73 | 51 | 51 | 69.86% | 69.86% |
| 5/19/2014 2:00 | Math-PT-Cell Phon | MARIONVILLE R-IX | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Cell Phon | MARIONVILLE R-IX | 1 | 58 | 46 | 46 | 79.31% | 79.31% |
| 5/19/2014 2:00 | Math-PT-Donuts | MARIONVILLE R-IX | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | MARIONVILLE R-IX | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARIONVILLE R-IX | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARIONVILLE R-IX | 1 | 30 | 24 | 24 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARIONVILLE R-IX | 1 | 28 | 22 | 22 | 78.57% | 78.57% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MARIONVILLE R-IX | 1 | 73 | 51 | 51 | 69.86% | 69.86% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | MARQUAND-ZION | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | MARQUAND-ZION | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MARQUAND-ZION | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MARQUAND-ZION | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-The Americ | MARSHALL | 1 | 185 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MARSHALL | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MARSHALL | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MARSHALL | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MARSHALL | 1 | 47 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Technology | MARSHFIELD R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | MARSHFIELD R-I | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | Math-PT-South Po | MARSHFIELD R-I | 1 | 241 | 231 | 231 | 95.85% | 95.85% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | MARSHFIELD R-I | 1 | 2 | 1 | 0 | 50.00% | 0.00% |

| Date | Test | District | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARSHFIELD R-I | 1 | 58 | 58 | 58 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARSHFIELD R-I | 1 | 58 | 58 | 58 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARSHFIELD R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARSHFIELD R-I | 1 | 63 | 57 | 57 | 90.48% | 90.48% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MARSHFIELD R-I | 1 | 62 | 59 | 59 | 95.16% | 95.16% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | MCDONALD CO. R- | 1 | 39 | 35 | 35 | 89.74% | 89.74% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MCDONALD CO. R- | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MCDONALD CO. R- | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MCDONALD CO. R- | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MCDONALD CO. R- | 1 | 9 | 8 | 8 | 88.89% | 88.89% |
| 5/19/2014 2:00 | Math-PT-Camping | MEADOW HEIGHT | 1 | 31 | 28 | 28 | 90.32% | 90.32% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MEADOW HEIGHT | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MEADOW HEIGHT | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MEADOW HEIGHT | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MEADOW HEIGHT | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | MEHLVILLE R-IX | 1 | 85 | 79 | 79 | 92.94% | 92.94% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MEHLVILLE R-IX | 1 | 17 | 14 | 13 | 82.35% | 76.47% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MEHLVILLE R-IX | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MEHLVILLE R-IX | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MEHLVILLE R-IX | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MEHLVILLE R-IX | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | MERAMEC VALLEY | 1 | 72 | 65 | 65 | 90.28% | 90.28% |
| 5/19/2014 2:00 | ELA-PT-Technology | MERAMEC VALLEY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Trees-A | MERAMEC VALLEY | 1 | 59 | 55 | 55 | 93.22% | 93.22% |
| 5/19/2014 2:00 | Math-PT-Camping | MERAMEC VALLEY | 1 | 214 | 199 | 196 | 92.99% | 91.59% |
| 5/19/2014 2:00 | Math-PT-Donuts | MERAMEC VALLEY | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MERAMEC VALLEY | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MERAMEC VALLEY | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MERAMEC VALLEY | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MERAMEC VALLEY | 1 | 12 | 11 | 11 | 91.67% | 91.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MERAMEC VALLEY | 1 | 11 | 9 | 9 | 81.82% | 81.82% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MERAMEC VALLEY | 1 | 18 | 18 | 16 | 100.00% | 88.89% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MERAMEC VALLEY | 1 | 18 | 15 | 15 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MERAMEC VALLEY | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | MERAMEC VALLEY | 1 | 18 | 17 | 17 | 94.44% | 94.44% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | MERAMEC VALLEY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MERAMEC VALLEY | 1 | 49 | 49 | 46 | 100.00% | 93.88% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MERAMEC VALLEY | 1 | 52 | 52 | 48 | 100.00% | 92.31% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MERAMEC VALLEY | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MERAMEC VALLEY | 1 | 57 | 52 | 51 | 91.23% | 89.47% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MERAMEC VALLEY | 1 | 56 | 51 | 49 | 91.07% | 87.50% |
| 5/19/2014 2:00 | ELA-PT-Technology | MEXICO 59 | 1 | 190 | 185 | 184 | 97.37% | 96.84% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | MEXICO 59 | 1 | 151 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great | MEXICO 59 | 1 | 175 | 151 | 151 | 86.29% | 86.29% |
| 5/19/2014 2:00 | Math-PT-Donuts | MEXICO 59 | 1 | 97 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | MEXICO 59 | 1 | 49 | 48 | 48 | 97.96% | 97.96% |
| 5/19/2014 2:00 | Math-PT-Making S | MEXICO 59 | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MEXICO 59 | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MEXICO 59 | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | MEXICO 59 | 1 | 49 | 48 | 48 | 97.96% | 97.96% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | MEXICO 59 | 1 | 190 | 187 | 187 | 98.42% | 98.42% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | MEXICO 59 | 1 | 97 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | MEXICO 59 | 1 | 151 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MEXICO 59 | 1 | 175 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | MILLER CO. R-III | 1 | 16 | 16 | 16 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MILLER CO. R-III | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MILLER CO. R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MILLER CO. R-III | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MILLER CO. R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MILLER CO. R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MILLER CO. R-III | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MILLER CO. R-III | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | HS-Math-PT-Great | MO SCHOOL FOR 1 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | MO SCHOOL FOR 1 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Baseball | MOBERLY | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Baseball- | MOBERLY | 1 | 171 | 131 | 131 | 76.61% | 76.61% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | MOBERLY | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | MOBERLY | 1 | 162 | 149 | 149 | 91.98% | 91.98% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | MOBERLY | 1 | 41 | 39 | 38 | 95.12% | 92.68% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | MOBERLY | 1 | 40 | 38 | 35 | 95.00% | 87.50% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | MOBERLY | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | MOBERLY | 1 | 41 | 37 | 37 | 90.24% | 90.24% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | MOBERLY | 1 | 40 | 36 | 36 | 90.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | MOBERLY | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | MOBERLY | 1 | 69 | 51 | 51 | 73.91% | 73.91% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | MOBERLY | 1 | 68 | 51 | 51 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | MOBERLY | 1 | 34 | 29 | 29 | 85.29% | 85.29% |
| 5/19/2014 2:00 | ELA-PT-Growth an | MONETT R-I | 1 | 166 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | MONETT R-I | 1 | 172 | 166 | 166 | 96.51% | 96.51% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | MONETT R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | MONETT R-I | 1 | 166 | 23 | 0 | 13.86% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MONETT R-I | 1 | 172 | 166 | 166 | 96.51% | 96.51% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | MONETT R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | MONITEAU CO. R-I | 1 | 92 | 86 | 86 | 93.48% | 93.48% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MONITEAU CO. R-I | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | MONITEAU CO. R-I | 1 | 18 | 16 | 16 | 88.89% | 88.89% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G04-ELA-Noi MONITEAU CO. R-I | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi MONITEAU CO. R-I | 1 | 18 | 17 | 17 | 94.44% | 94.44% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi MONITEAU CO. R-I | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Technolog MONROE CITY R-I | 1 | 50 | 7 | 4 | 14.00% | 8.00% |
| 5/19/2014 2:00 SBAC-G06-ELA-Noi MONROE CITY R-I | 1 | 13 | 13 | 9 | 100.00% | 69.23% |
| 5/19/2014 2:00 SBAC-G06-ELA-Noi MONROE CITY R-I | 1 | 12 | 10 | 6 | 83.33% | 50.00% |
| 5/19/2014 2:00 SBAC-G06-ELA-Noi MONROE CITY R-I | 1 | 13 | 13 | 7 | 100.00% | 53.85% |
| 5/19/2014 2:00 SBAC-G06-ELA-Noi MONROE CITY R-I | 1 | 12 | 10 | 7 | 83.33% | 58.33% |
| 5/19/2014 2:00 HS-Math-PT-Great MONTROSE R-XIV | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-Math-No MONTROSE R-XIV | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Sandbox- MORGAN CO. R-II | 1 | 112 | 106 | 106 | 94.64% | 94.64% |
| 5/19/2014 2:00 SBAC-G06-Math-N MORGAN CO. R-II | 1 | 56 | 53 | 53 | 94.64% | 94.64% |
| 5/19/2014 2:00 SBAC-G06-Math-N MORGAN CO. R-II | 1 | 56 | 53 | 53 | 94.64% | 94.64% |
| 5/19/2014 2:00 HS-Math-PT-Great NEELYVILLE R-IV | 1 | 42 | 33 | 33 | 78.57% | 78.57% |
| 5/19/2014 2:00 Math-PT-School Li NEELYVILLE R-IV | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G03-Math-N NEELYVILLE R-IV | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-HS-Math-No NEELYVILLE R-IV | 1 | 42 | 33 | 33 | 78.57% | 78.57% |
| 5/19/2014 2:00 ELA-PT-Marine Ani NELL HOLCOMB R- | 1 | 27 | 26 | 26 | 96.30% | 96.30% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NELL HOLCOMB R- | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NELL HOLCOMB R- | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NELL HOLCOMB R- | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NELL HOLCOMB R- | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NELL HOLCOMB R- | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Renewable NEOSHO R-V | 1 | 114 | 100 | 97 | 87.72% | 85.09% |
| 5/19/2014 2:00 ELA-PT-Uncommoi NEOSHO R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-Math-PT-Great NEOSHO R-V | 1 | 37 | 32 | 32 | 86.49% | 86.49% |
| 5/19/2014 2:00 Math-PT-Animal Ju NEOSHO R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Animal Ju NEOSHO R-V | 1 | 111 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-Noi NEOSHO R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 22 | 20 | 20 | 90.91% | 90.91% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 22 | 19 | 19 | 86.36% | 86.36% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 23 | 21 | 21 | 91.30% | 91.30% |
| 5/19/2014 2:00 SBAC-G04-Math-N NEOSHO R-V | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 SBAC-G08-ELA-Noi NEOSHO R-V | 1 | 23 | 21 | 20 | 91.30% | 86.96% |
| 5/19/2014 2:00 SBAC-G08-ELA-Noi NEOSHO R-V | 1 | 23 | 23 | 21 | 100.00% | 91.30% |
| 5/19/2014 2:00 SBAC-G08-ELA-Noi NEOSHO R-V | 1 | 23 | 20 | 18 | 86.96% | 78.26% |
| 5/19/2014 2:00 SBAC-G08-ELA-Noi NEOSHO R-V | 1 | 23 | 23 | 22 | 100.00% | 95.65% |
| 5/19/2014 2:00 SBAC-G08-ELA-Noi NEOSHO R-V | 1 | 22 | 20 | 19 | 90.91% | 86.36% |
| 5/19/2014 2:00 SBAC-HS-Math-No NEOSHO R-V | 1 | 37 | 32 | 32 | 86.49% | 86.49% |
| 5/19/2014 2:00 HS-ELA-PT-A New NEVADA R-V | 1 | 1 | 1 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 HS-ELA-PT-A New NEVADA R-V | 1 | 204 | 0 | 0 | 0.00% | 0.00% |

| Date | Test | District | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | Math-PT-Commun | NEVADA R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Commun | NEVADA R-V | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | NEVADA R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | NEVADA R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | NEVADA R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NEVADA R-V | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NEVADA R-V | 1 | 122 | 92 | 75 | 75.41% | 61.48% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NEVADA R-V | 1 | 82 | 62 | 50 | 75.61% | 60.98% |
| 5/19/2014 2:00 | ELA-PT-Technolog | NEW MADRID CO. | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | NEW MADRID CO. | 1 | 51 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | NEW MADRID CO. | 1 | 51 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | NEW MADRID CO. | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | NIANGUA R-V | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | NIANGUA R-V | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | NIANGUA R-V | 1 | 16 | 13 | 13 | 81.25% | 81.25% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | NIANGUA R-V | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | NIANGUA R-V | 1 | 8 | 7 | 6 | 87.50% | 75.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | NIANGUA R-V | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | NIANGUA R-V | 1 | 9 | 9 | 8 | 100.00% | 88.89% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NIANGUA R-V | 1 | 16 | 16 | 16 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NIANGUA R-V | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | NIANGUA R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Importanc | NORBORNE R-VIII | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | NORBORNE R-VIII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | NORBORNE R-VIII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | NORBORNE R-VIII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | NORBORNE R-VIII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | NORTH CALLAWAY | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | NORTH CALLAWAY | 1 | 34 | 33 | 32 | 97.06% | 94.12% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | NORTH CALLAWAY | 1 | 92 | 75 | 75 | 81.52% | 81.52% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | NORTH CALLAWAY | 1 | 20 | 20 | 20 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | NORTH CALLAWAY | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | NORTH CALLAWAY | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 7 | 7 | 6 | 100.00% | 85.71% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 7 | 7 | 4 | 100.00% | 57.14% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 7 | 7 | 5 | 100.00% | 71.43% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | NORTH CALLAWAY | 1 | 6 | 6 | 3 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | NORTH CALLAWAY | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | NORTH CALLAWAY | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | NORTH CALLAWAY | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | NORTH CALLAWAY | 1 | 5 | 5 | 5 | 100.00% | 100.00% |

| | | | | | |
|---|---|---|---|---|---|
| 5/19/2014 2:00 SBAC-G08-ELA-No NORTH CALLAWAY | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-HS-ELA-Nonl NORTH CALLAWAY | 1 | 92 | 75 | 48 | 81.52% | 52.17% |
| 5/19/2014 2:00 HS-Math-PT-Great NORTH HARRISON | 1 | 22 | 19 | 19 | 86.36% | 86.36% |
| 5/19/2014 2:00 SBAC-HS-Math-No NORTH HARRISON | 1 | 22 | 19 | 19 | 86.36% | 86.36% |
| 5/19/2014 2:00 HS-Math-PT-Great NORTH MERCER C | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-HS-Math-No NORTH MERCER C | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 Math-PT-Sandbox- NORTH NODAWAY | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N NORTH NODAWAY | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N NORTH NODAWAY | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N NORTH NODAWAY | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N NORTH NODAWAY | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 Math-PT-South Po NORTH PLATTE CO | 1 | 33 | 33 | 33 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N NORTH PLATTE CO | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N NORTH PLATTE CO | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G08-Math-N NORTH PLATTE CO | 1 | 6 | 5 | 5 | 83.33% | 83.33% |
| 5/19/2014 2:00 SBAC-G08-Math-N NORTH PLATTE CO | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 SBAC-G08-Math-N NORTH PLATTE CO | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 Math-PT-Donuts-A NORTHEAST VERN | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G07-Math-N NORTHEAST VERN | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G07-Math-N NORTHEAST VERN | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Deserts NORTHWEST R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 ELA-PT-Deserts-A NORTHWEST R-I | 1 | 70 | 65 | 63 | 92.86% | 90.00% |
| 5/19/2014 2:00 ELA-PT-Trees NORTHWEST R-I | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 ELA-PT-Trees-A NORTHWEST R-I | 1 | 81 | 79 | 79 | 97.53% | 97.53% |
| 5/19/2014 2:00 HS-ELA-PT-A New NORTHWEST R-I | 1 | 514 | 443 | 441 | 86.19% | 85.80% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 30 | 28 | 28 | 93.33% | 93.33% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 30 | 30 | 30 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 30 | 29 | 29 | 96.67% | 96.67% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 30 | 30 | 29 | 100.00% | 96.67% |
| 5/19/2014 2:00 SBAC-G04-ELA-No NORTHWEST R-I | 1 | 31 | 29 | 29 | 93.55% | 93.55% |
| 5/19/2014 2:00 SBAC-HS-ELA-Nonl NORTHWEST R-I | 1 | 514 | 445 | 443 | 86.58% | 86.19% |
| 5/19/2014 2:00 ELA-PT-Trees-A OAK GROVE R-VI | 1 | 150 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G03-ELA-No OAK GROVE R-VI | 1 | 50 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G03-ELA-No OAK GROVE R-VI | 1 | 51 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 SBAC-G03-ELA-No OAK GROVE R-VI | 1 | 49 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 HS-Math-PT-Great OAK RIDGE R-VI | 1 | 26 | 24 | 24 | 92.31% | 92.31% |
| 5/19/2014 2:00 SBAC-HS-Math-No OAK RIDGE R-VI | 1 | 26 | 25 | 25 | 96.15% | 96.15% |
| 5/19/2014 2:00 HS-ELA-PT-A New ODESSA R-VII | 1 | 173 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 Math-PT-Talent Sh ODESSA R-VII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 Math-PT-Talent Sh ODESSA R-VII | 1 | 169 | 162 | 162 | 95.86% | 95.86% |
| 5/19/2014 2:00 SBAC-G06-Math-N ODESSA R-VII | 1 | 19 | 19 | 19 | 100.00% | 100.00% |
| 5/19/2014 2:00 SBAC-G06-Math-N ODESSA R-VII | 1 | 19 | 19 | 19 | 100.00% | 100.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G06-Math-N ODESSA R-VII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N ODESSA R-VII | 1 | 65 | 62 | 62 | 95.38% | 95.38% |
| 5/19/2014 2:00 | SBAC-G06-Math-N ODESSA R-VII | 1 | 66 | 62 | 62 | 93.94% | 93.94% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl ODESSA R-VII | 1 | 100 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl ODESSA R-VII | 1 | 73 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Land Form ORCHARD FARM R | 1 | 131 | 118 | 113 | 90.08% | 86.26% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nol ORCHARD FARM R | 1 | 131 | 119 | 117 | 90.84% | 89.31% |
| 5/19/2014 2:00 | HS-Math-PT-Great OSAGE CO. R-I | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No OSAGE CO. R-I | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an OSAGE CO. R-II | 1 | 43 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great OSAGE CO. R-II | 1 | 45 | 45 | 45 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh OSAGE CO. R-II | 1 | 43 | 43 | 43 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nol OSAGE CO. R-II | 1 | 43 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N OSAGE CO. R-II | 1 | 43 | 43 | 43 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No OSAGE CO. R-II | 1 | 45 | 45 | 45 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great OSCEOLA | 1 | 38 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No OSCEOLA | 1 | 38 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable OTTERVILLE R-VI | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo OTTERVILLE R-VI | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N OTTERVILLE R-VI | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N OTTERVILLE R-VI | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N OTTERVILLE R-VI | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol OTTERVILLE R-VI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol OTTERVILLE R-VI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol OTTERVILLE R-VI | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol OTTERVILLE R-VI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol OTTERVILLE R-VI | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New  PARIS R-II | 1 | 26 | 26 | 26 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great PARIS R-II | 1 | 33 | 32 | 32 | 96.97% | 96.97% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol PARIS R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol PARIS R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol PARIS R-II | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol PARIS R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol PARIS R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No PARIS R-II | 1 | 32 | 31 | 31 | 96.88% | 96.88% |
| 5/19/2014 2:00 | ELA-PT-Growth an PARK HILL | 1 | 65 | 51 | 40 | 78.46% | 61.54% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol PARK HILL | 1 | 16 | 15 | 13 | 93.75% | 81.25% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol PARK HILL | 1 | 16 | 16 | 12 | 100.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol PARK HILL | 1 | 16 | 14 | 14 | 87.50% | 87.50% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol PARK HILL | 1 | 17 | 17 | 15 | 100.00% | 88.24% |
| 5/19/2014 2:00 | Math-PT-Order Fo PARKWAY C-2 | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | Math-PT-Order Fo PARKWAY C-2 | 1 | 75 | 69 | 66 | 92.00% | 88.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N PARKWAY C-2 | 1 | 25 | 24 | 23 | 96.00% | 92.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G03-Math-N | PARKWAY C-2 | 1 | 25 | 24 | 20 | 96.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | PARKWAY C-2 | 1 | 25 | 21 | 21 | 84.00% | 84.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | PARKWAY C-2 | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | PATTONSBURG R-I | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | PATTONSBURG R-I | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Camping | PATTONVILLE R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Camping | PATTONVILLE R-III | 1 | 185 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Commun | PATTONVILLE R-III | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Commun | PATTONVILLE R-III | 1 | 118 | 51 | 50 | 43.22% | 42.37% |
| 5/19/2014 2:00 | Math-PT-Sandbox | PATTONVILLE R-III | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | PATTONVILLE R-III | 1 | 174 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | PATTONVILLE R-III | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | PATTONVILLE R-III | 1 | 68 | 59 | 56 | 86.76% | 82.35% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 23 | 7 | 7 | 30.43% | 30.43% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 23 | 11 | 10 | 47.83% | 43.48% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 25 | 14 | 13 | 56.00% | 52.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 23 | 9 | 9 | 39.13% | 39.13% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | PATTONVILLE R-III | 1 | 24 | 11 | 10 | 45.83% | 41.67% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | PATTONVILLE R-III | 1 | 17 | 16 | 15 | 94.12% | 88.24% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | PATTONVILLE R-III | 1 | 17 | 15 | 11 | 88.24% | 64.71% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | PATTONVILLE R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | PATTONVILLE R-III | 1 | 17 | 15 | 13 | 88.24% | 76.47% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | PATTONVILLE R-III | 1 | 17 | 16 | 12 | 94.12% | 70.59% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | PATTONVILLE R-III | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | PATTONVILLE R-III | 1 | 86 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | PATTONVILLE R-III | 1 | 88 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PATTONVILLE R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PATTONVILLE R-III | 1 | 93 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PATTONVILLE R-III | 1 | 92 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Space Expl | PERRY CO. 32 | 1 | 199 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PERRY CO. 32 | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PERRY CO. 32 | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PERRY CO. 32 | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PERRY CO. 32 | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PERRY CO. 32 | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | PETTIS CO. R-V | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PETTIS CO. R-V | 1 | 4 | 3 | 3 | 75.00% | 75.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PETTIS CO. R-V | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PETTIS CO. R-V | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PETTIS CO. R-V | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | PETTIS CO. R-V | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | PETTIS CO. R-XII | 1 | 14 | 13 | 12 | 92.86% | 85.71% |

| Date | Assessment | District | | | | | |
|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G07-Math-N | PETTIS CO. R-XII | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PETTIS CO. R-XII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PETTIS CO. R-XII | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PETTIS CO. R-XII | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PIERCE CITY R-VI | 1 | 58 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PIERCE CITY R-VI | 1 | 58 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | PIKE CO. R-III | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | PIKE CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PIKE CO. R-III | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PIKE CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PILOT GROVE C-4 | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PILOT GROVE C-4 | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | ELA-PT-Importance | PLATTE CO. R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Importance | PLATTE CO. R-III | 1 | 83 | 80 | 80 | 96.39% | 96.39% |
| 5/19/2014 2:00 | ELA-PT-Technology | PLATTE CO. R-III | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | PLATTE CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | PLATTE CO. R-III | 1 | 20 | 20 | 18 | 100.00% | 90.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | PLATTE CO. R-III | 1 | 21 | 21 | 21 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | PLATTE CO. R-III | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | PLATTE CO. R-III | 1 | 21 | 21 | 20 | 100.00% | 95.24% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | PLATTE CO. R-III | 1 | 21 | 20 | 20 | 95.24% | 95.24% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | PLATTE CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PLEASANT HILL R-I | 1 | 174 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PLEASANT HILL R-I | 1 | 174 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PLEASANT HOPE R | 1 | 52 | 49 | 49 | 94.23% | 94.23% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PLEASANT HOPE R | 1 | 52 | 48 | 48 | 92.31% | 92.31% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | PRAIRIE HOME R-V | 1 | 11 | 11 | 11 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | PRAIRIE HOME R-V | 1 | 11 | 11 | 11 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Making S | PRINCETON R-V | 1 | 26 | 26 | 26 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | PRINCETON R-V | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | PRINCETON R-V | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | PRINCETON R-V | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PUTNAM CO. R-I | 1 | 50 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PUTNAM CO. R-I | 1 | 50 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | PUXICO R-VIII | 1 | 54 | 39 | 39 | 72.22% | 72.22% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | PUXICO R-VIII | 1 | 54 | 50 | 48 | 92.59% | 88.89% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | REEDS SPRING R-IV | 1 | 148 | 121 | 115 | 81.76% | 77.70% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Non | REEDS SPRING R-IV | 1 | 148 | 130 | 127 | 87.84% | 85.81% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | RICHARDS R-V | 1 | 42 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | RICHARDS R-V | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | RICHARDS R-V | 1 | 42 | 17 | 0 | 40.48% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | RICHARDS R-V | 1 | 1 | 1 | 0 | 100.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | RICHLAND R-I | 1 | 15 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great | RICHLAND R-I | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | RICHLAND R-I | 1 | 15 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | RICHLAND R-I | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | RICHMOND R-XVI | 1 | 106 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | RICHMOND R-XVI | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | RICHMOND R-XVI | 1 | 30 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | RICHMOND R-XVI | 1 | 89 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | RICHMOND R-XVI | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | RICHMOND R-XVI | 1 | 106 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | RICHMOND R-XVI | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | RICHMOND R-XVI | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | RICHMOND R-XVI | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | RICHMOND R-XVI | 1 | 30 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | RICHMOND R-XVI | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | RICHMOND R-XVI | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | RITENOUR | 1 | 66 | 62 | 61 | 93.94% | 92.42% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | RITENOUR | 1 | 22 | 21 | 19 | 95.45% | 86.36% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | RITENOUR | 1 | 23 | 22 | 21 | 95.65% | 91.30% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | RITENOUR | 1 | 21 | 21 | 18 | 100.00% | 85.71% |
| 5/19/2014 2:00 | Math-PT-Order Fo | ROLLA 31 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Order Fo | ROLLA 31 | 1 | 103 | 94 | 94 | 91.26% | 91.26% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ROLLA 31 | 1 | 35 | 33 | 33 | 94.29% | 94.29% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ROLLA 31 | 1 | 34 | 34 | 33 | 100.00% | 97.06% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ROLLA 31 | 1 | 34 | 31 | 30 | 91.18% | 88.24% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ROLLA 31 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Aztec Emp | SALISBURY R-IV | 1 | 34 | 34 | 34 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SALISBURY R-IV | 1 | 37 | 37 | 37 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | SALISBURY R-IV | 1 | 34 | 34 | 34 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SALISBURY R-IV | 1 | 37 | 37 | 37 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SARCOXIE R-II | 1 | 52 | 36 | 36 | 69.23% | 69.23% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SARCOXIE R-II | 1 | 52 | 36 | 36 | 69.23% | 69.23% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | SAVANNAH R-III | 1 | 169 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Camping | SAVANNAH R-III | 1 | 204 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SAVANNAH R-III | 1 | 82 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SAVANNAH R-III | 1 | 82 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SAVANNAH R-III | 1 | 40 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | SAVANNAH R-III | 1 | 101 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | SAVANNAH R-III | 1 | 68 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SCHUYLER CO. R-I | 1 | 46 | 36 | 36 | 78.26% | 78.26% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SCHUYLER CO. R-I | 1 | 46 | 37 | 37 | 80.43% | 80.43% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | SCOTLAND CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | SCOTLAND CO. R-I | 1 | 38 | 38 | 38 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | SCOTLAND CO. R-I | 1 | 10 | 10 | 10 | 100.00% | 100.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G05-Math-N | SCOTLAND CO. R-I | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | SCOTLAND CO. R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | SCOTLAND CO. R-I | 1 | 10 | 10 | 10 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | SCOTLAND CO. R-I | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | SCOTT CO. R-IV | 1 | 61 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | SCOTT CO. R-IV | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | SCOTT CO. R-IV | 1 | 22 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SENATH-HORNERS | 1 | 52 | 36 | 36 | 69.23% | 69.23% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SENATH-HORNERS | 1 | 52 | 36 | 36 | 69.23% | 69.23% |
| 5/19/2014 2:00 | Math-PT-South Po | SHAWNEE R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHAWNEE R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | SHELBY CO. R-IV | 1 | 55 | 54 | 54 | 98.18% | 98.18% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | SHELBY CO. R-IV | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | SHELBY CO. R-IV | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | SHELBY CO. R-IV | 1 | 37 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Jt | SHELBY CO. R-IV | 1 | 37 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | SHELBY CO. R-IV | 1 | 55 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | SHELBY CO. R-IV | 1 | 39 | 37 | 34 | 94.87% | 87.18% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | SHELBY CO. R-IV | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | SHELBY CO. R-IV | 1 | 37 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHELBY CO. R-IV | 1 | 55 | 54 | 54 | 98.18% | 98.18% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELBY CO. R-IV | 1 | 55 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | SHELDON R-VIII | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHELDON R-VIII | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | Math-PT-Baseball- | SHELL KNOB 78 | 1 | 6 | 4 | 4 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELL KNOB 78 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELL KNOB 78 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELL KNOB 78 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELL KNOB 78 | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SHELL KNOB 78 | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Space Expl | SHERWOOD CASS | 1 | 60 | 57 | 57 | 95.00% | 95.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHERWOOD CASS | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHERWOOD CASS | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHERWOOD CASS | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHERWOOD CASS | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | SHERWOOD CASS | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | SIKESTON R-6 | 1 | 256 | 194 | 194 | 75.78% | 75.78% |
| 5/19/2014 2:00 | Math-PT-Making S | SIKESTON R-6 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | SIKESTON R-6 | 1 | 78 | 75 | 75 | 96.15% | 96.15% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SIKESTON R-6 | 1 | 25 | 24 | 24 | 96.00% | 96.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SIKESTON R-6 | 1 | 26 | 26 | 26 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SIKESTON R-6 | 1 | 27 | 25 | 25 | 92.59% | 92.59% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SIKESTON R-6 | 1 | 1 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-HS-ELA-Non| SIKESTON R-6 | 1 | 256 | 199 | 195 | 77.73% | 76.17% |
| 5/19/2014 2:00 | ELA-PT-Animals W | SILEX R-I | 1 | 29 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An| SILEX R-I | 1 | 28 | 27 | 20 | 96.43% | 71.43% |
| 5/19/2014 2:00 | Math-PT-Donuts-A| SILEX R-I | 1 | 23 | 20 | 20 | 86.96% | 86.96% |
| 5/19/2014 2:00 | Math-PT-Making S | SILEX R-I | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | SILEX R-I | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No| SILEX R-I | 1 | 29 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SILEX R-I | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SILEX R-I | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | SILEX R-I | 1 | 21 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No| SILEX R-I | 1 | 28 | 27 | 26 | 96.43% | 92.86% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | SILEX R-I | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | SILEX R-I | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | SKYLINE R-II | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-South Po | SKYLINE R-II | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SKYLINE R-II | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SKYLINE R-II | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | SOUTH CALLAWAY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | SOUTH CALLAWAY | 1 | 61 | 42 | 41 | 68.85% | 67.21% |
| 5/19/2014 2:00 | Math-PT-South Po | SOUTH CALLAWAY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 11 | 11 | 11 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 12 | 8 | 8 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No| SOUTH CALLAWAY | 1 | 13 | 12 | 12 | 92.31% | 92.31% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | SOUTH CALLAWAY | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SOUTH HARRISON | 1 | 47 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SOUTH HARRISON | 1 | 47 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo| SOUTH PEMISCOT | 1 | 37 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo| SOUTH PEMISCOT | 1 | 6 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | SOUTH PEMISCOT | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | SOUTH PEMISCOT | 1 | 31 | 28 | 28 | 90.32% | 90.32% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No| SOUTH PEMISCOT | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No| SOUTH PEMISCOT | 1 | 37 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No| SOUTH PEMISCOT | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | SOUTH PEMISCOT | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | SOUTH PEMISCOT | 1 | 12 | 10 | 10 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | SOUTH PEMISCOT | 1 | 16 | 14 | 14 | 87.50% | 87.50% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | SOUTH PEMISCOT | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | ELA-PT-Growth an | SPECL. SCH. DST. S | 1 | 7 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | SPECL. SCH. DST. S | 1 | 14 | 4 | 4 | 28.57% | 28.57% |
| 5/19/2014 2:00 | HS-ELA-PT-A New  | SPECL. SCH. DST. S | 1 | 178 | 163 | 143 | 91.57% | 80.34% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | HS-Math-PT-Great SPECL. SCH. DST. S | 1 | 41 | 6 | 6 | 14.63% | 14.63% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor SPECL. SCH. DST. S | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor SPECL. SCH. DST. S | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor SPECL. SCH. DST. S | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor SPECL. SCH. DST. S | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor SPECL. SCH. DST. S | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor SPECL. SCH. DST. S | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor SPECL. SCH. DST. S | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor SPECL. SCH. DST. S | 1 | 3 | 1 | 1 | 33.33% | 33.33% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor SPECL. SCH. DST. S | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI SPECL. SCH. DST. S | 1 | 105 | 95 | 57 | 90.48% | 54.29% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI SPECL. SCH. DST. S | 1 | 73 | 66 | 46 | 90.41% | 63.01% |
| 5/19/2014 2:00 | SBAC-HS-Math-No SPECL. SCH. DST. S | 1 | 21 | 6 | 6 | 28.57% | 28.57% |
| 5/19/2014 2:00 | SBAC-HS-Math-No SPECL. SCH. DST. S | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-The Ameri SPRINGFIELD R-XII | 1 | 30 | 21 | 21 | 70.00% | 70.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo SPRINGFIELD R-XII | 1 | 55 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New SPRINGFIELD R-XII | 1 | 369 | 345 | 335 | 93.50% | 90.79% |
| 5/19/2014 2:00 | HS-ELA-PT-A New SPRINGFIELD R-XII | 1 | 462 | 213 | 208 | 46.10% | 45.02% |
| 5/19/2014 2:00 | HS-Math-PT-Great SPRINGFIELD R-XII | 1 | 368 | 277 | 277 | 75.27% | 75.27% |
| 5/19/2014 2:00 | Math-PT-Making S SPRINGFIELD R-XII | 1 | 55 | 52 | 51 | 94.55% | 92.73% |
| 5/19/2014 2:00 | Math-PT-Order Fo SPRINGFIELD R-XII | 1 | 79 | 72 | 71 | 91.14% | 89.87% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha SPRINGFIELD R-XII | 1 | 60 | 50 | 50 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor SPRINGFIELD R-XII | 1 | 18 | 17 | 16 | 94.44% | 88.89% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor SPRINGFIELD R-XII | 1 | 18 | 16 | 16 | 88.89% | 88.89% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor SPRINGFIELD R-XII | 1 | 19 | 15 | 14 | 78.95% | 73.68% |
| 5/19/2014 2:00 | SBAC-G03-Math-N SPRINGFIELD R-XII | 1 | 33 | 32 | 32 | 96.97% | 96.97% |
| 5/19/2014 2:00 | SBAC-G03-Math-N SPRINGFIELD R-XII | 1 | 32 | 30 | 30 | 93.75% | 93.75% |
| 5/19/2014 2:00 | SBAC-G03-Math-N SPRINGFIELD R-XII | 1 | 31 | 29 | 27 | 93.55% | 87.10% |
| 5/19/2014 2:00 | SBAC-G04-Math-N SPRINGFIELD R-XII | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N SPRINGFIELD R-XII | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N SPRINGFIELD R-XII | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N SPRINGFIELD R-XII | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N SPRINGFIELD R-XII | 1 | 8 | 8 | 7 | 100.00% | 87.50% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor SPRINGFIELD R-XII | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor SPRINGFIELD R-XII | 1 | 8 | 5 | 5 | 62.50% | 62.50% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor SPRINGFIELD R-XII | 1 | 7 | 5 | 5 | 71.43% | 71.43% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor SPRINGFIELD R-XII | 1 | 8 | 5 | 5 | 62.50% | 62.50% |
| 5/19/2014 2:00 | SBAC-G05-Math-N SPRINGFIELD R-XII | 1 | 15 | 9 | 9 | 60.00% | 60.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N SPRINGFIELD R-XII | 1 | 15 | 14 | 14 | 93.33% | 93.33% |
| 5/19/2014 2:00 | SBAC-G05-Math-N SPRINGFIELD R-XII | 1 | 15 | 15 | 15 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N SPRINGFIELD R-XII | 1 | 15 | 11 | 11 | 73.33% | 73.33% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI SPRINGFIELD R-XII | 1 | 369 | 360 | 327 | 97.56% | 88.62% |
| 5/19/2014 2:00 | SBAC-HS-ELA-NonI SPRINGFIELD R-XII | 1 | 267 | 150 | 132 | 56.18% | 49.44% |

| Date | Test | District | | Count 1 | Count 2 | Count 3 | Pct 1 | Pct 2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | SPRINGFIELD R-XII | 1 | 195 | 102 | 98 | 52.31% | 50.26% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SPRINGFIELD R-XII | 1 | 368 | 285 | 280 | 77.45% | 76.09% |
| 5/19/2014 2:00 | Math-PT-Commun | ST LOUIS LANG IM | 1 | 46 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | ST LOUIS LANG IM | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | ST LOUIS LANG IM | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | ST LOUIS LANG IM | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | ST LOUIS LANG IM | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | ST. CHARLES R-VI | 1 | 104 | 99 | 99 | 95.19% | 95.19% |
| 5/19/2014 2:00 | ELA-PT-Technolog | ST. CHARLES R-VI | 1 | 374 | 12 | 11 | 3.21% | 2.94% |
| 5/19/2014 2:00 | Math-PT-Animal Jt | ST. CHARLES R-VI | 1 | 29 | 29 | 29 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Making S | ST. CHARLES R-VI | 1 | 65 | 55 | 55 | 84.62% | 84.62% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | ST. CHARLES R-VI | 1 | 104 | 100 | 100 | 96.15% | 96.15% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. CHARLES R-VI | 1 | 31 | 27 | 25 | 87.10% | 80.65% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. CHARLES R-VI | 1 | 32 | 28 | 25 | 87.50% | 78.13% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. CHARLES R-VI | 1 | 31 | 30 | 27 | 96.77% | 87.10% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. CHARLES R-VI | 1 | 94 | 5 | 3 | 5.32% | 3.19% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. CHARLES R-VI | 1 | 93 | 5 | 3 | 5.38% | 3.23% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. CHARLES R-VI | 1 | 94 | 1 | 1 | 1.06% | 1.06% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. CHARLES R-VI | 1 | 93 | 1 | 0 | 1.08% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | ST. JAMES R-I | 1 | 116 | 95 | 94 | 81.90% | 81.03% |
| 5/19/2014 2:00 | Math-PT-Animal Jt | ST. JAMES R-I | 1 | 123 | 105 | 105 | 85.37% | 85.37% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. JAMES R-I | 1 | 40 | 34 | 34 | 85.00% | 85.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. JAMES R-I | 1 | 42 | 41 | 41 | 97.62% | 97.62% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. JAMES R-I | 1 | 41 | 34 | 32 | 82.93% | 78.05% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | ST. JAMES R-I | 1 | 70 | 54 | 52 | 77.14% | 74.29% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | ST. JAMES R-I | 1 | 46 | 40 | 38 | 86.96% | 82.61% |
| 5/19/2014 2:00 | ELA-PT-Heatwaves | ST. JOSEPH | 1 | 39 | 38 | 37 | 97.44% | 94.87% |
| 5/19/2014 2:00 | ELA-PT-Importanc | ST. JOSEPH | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Importanc | ST. JOSEPH | 1 | 233 | 30 | 30 | 12.88% | 12.88% |
| 5/19/2014 2:00 | ELA-PT-The Ameri | ST. JOSEPH | 1 | 67 | 59 | 58 | 88.06% | 86.57% |
| 5/19/2014 2:00 | ELA-PT-Trees | ST. JOSEPH | 1 | 67 | 61 | 61 | 91.04% | 91.04% |
| 5/19/2014 2:00 | Math-PT-Cell Phoi | ST. JOSEPH | 1 | 263 | 217 | 217 | 82.51% | 82.51% |
| 5/19/2014 2:00 | Math-PT-Science k | ST. JOSEPH | 1 | 46 | 45 | 43 | 97.83% | 93.48% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Noi | ST. JOSEPH | 1 | 106 | 98 | 94 | 92.45% | 88.68% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | ST. JOSEPH | 1 | 46 | 45 | 44 | 97.83% | 95.65% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Noi | ST. JOSEPH | 1 | 17 | 17 | 15 | 100.00% | 88.24% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Noi | ST. JOSEPH | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Noi | ST. JOSEPH | 1 | 17 | 17 | 16 | 100.00% | 94.12% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Noi | ST. JOSEPH | 1 | 16 | 16 | 15 | 100.00% | 93.75% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. JOSEPH | 1 | 58 | 4 | 4 | 6.90% | 6.90% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. JOSEPH | 1 | 58 | 10 | 10 | 17.24% | 17.24% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Noi | ST. JOSEPH | 1 | 3 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | ST. JOSEPH | 1 | 58 | 6 | 6 | 10.34% | 10.34% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | ST. JOSEPH | 1 | 59 | 10 | 10 | 16.95% | 16.95% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | ST. JOSEPH | 1 | 60 | 60 | 60 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | ST. JOSEPH | 1 | 51 | 51 | 51 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | ST. JOSEPH | 1 | 72 | 56 | 56 | 77.78% | 77.78% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | ST. JOSEPH | 1 | 80 | 53 | 53 | 66.25% | 66.25% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | ST. LOUIS CHARTEI | 1 | 49 | 48 | 46 | 97.96% | 93.88% |
| 5/19/2014 2:00 | ELA-PT-Marine Ani | ST. LOUIS CHARTEI | 1 | 49 | 48 | 48 | 97.96% | 97.96% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | ST. LOUIS CHARTEI | 1 | 13 | 13 | 13 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | ST. LOUIS CHARTEI | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | ST. LOUIS CHARTEI | 1 | 49 | 48 | 48 | 97.96% | 97.96% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | ST. LOUIS CHARTEI | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | ST. LOUIS CHARTEI | 1 | 9 | 9 | 9 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | ST. LOUIS CITY | 1 | 110 | 94 | 93 | 85.45% | 84.55% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | ST. LOUIS CITY | 1 | 110 | 77 | 68 | 70.00% | 61.82% |
| 5/19/2014 2:00 | ELA-PT-Growth an | STE. GENEVIEVE C( | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | STE. GENEVIEVE C( | 1 | 108 | 103 | 102 | 95.37% | 94.44% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | STE. GENEVIEVE C( | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | STE. GENEVIEVE C( | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox | STE. GENEVIEVE C( | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | STE. GENEVIEVE C( | 1 | 127 | 120 | 120 | 94.49% | 94.49% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | STE. GENEVIEVE C( | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-Nor | STE. GENEVIEVE C( | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | STE. GENEVIEVE C( | 1 | 33 | 33 | 33 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | STE. GENEVIEVE C( | 1 | 29 | 29 | 29 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | STE. GENEVIEVE C( | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | STE. GENEVIEVE C( | 1 | 29 | 25 | 25 | 86.21% | 86.21% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | STE. GENEVIEVE C( | 1 | 36 | 32 | 32 | 88.89% | 88.89% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | STE. GENEVIEVE C( | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | STE. GENEVIEVE C( | 1 | 107 | 106 | 105 | 99.07% | 98.13% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | STE. GENEVIEVE C( | 1 | 13 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | STE. GENEVIEVE C( | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | STE. GENEVIEVE C( | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | STEELVILLE R-III | 1 | 69 | 61 | 61 | 88.41% | 88.41% |
| 5/19/2014 2:00 | Math-PT-South Po | STEELVILLE R-III | 1 | 69 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nor | STEELVILLE R-III | 1 | 69 | 58 | 58 | 84.06% | 84.06% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | STEELVILLE R-III | 1 | 69 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Technolog | STOCKTON R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | STOCKTON R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | STOCKTON R-I | 1 | 85 | 84 | 84 | 98.82% | 98.82% |
| 5/19/2014 2:00 | SBAC-G07-ELA-Nor | STOCKTON R-I | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | STOCKTON R-I | 1 | 20 | 20 | 20 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | STOCKTON R-I | 1 | 21 | 21 | 21 | 100.00% | 100.00% |

| Date | Assessment | District | | Count | Val1 | Val2 | Pct1 | Pct2 |
|------|-----------|----------|---|-------|------|------|------|------|
| 5/19/2014 2:00 | SBAC-G07-Math-N | STOCKTON R-I | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | STOCKTON R-I | 1 | 22 | 21 | 21 | 95.45% | 95.45% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | STOCKTON R-I | 1 | 22 | 22 | 22 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | STOUTLAND R-II | 1 | 31 | 30 | 30 | 96.77% | 96.77% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | STOUTLAND R-II | 1 | 31 | 30 | 30 | 96.77% | 96.77% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | STRASBURG C-3 | 1 | 17 | 15 | 15 | 88.24% | 88.24% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | STRASBURG C-3 | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | STRASBURG C-3 | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | STRASBURG C-3 | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | STURGEON R-V | 1 | 28 | 28 | 28 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | STURGEON R-V | 1 | 36 | 31 | 31 | 86.11% | 86.11% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | STURGEON R-V | 1 | 28 | 28 | 20 | 100.00% | 71.43% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | STURGEON R-V | 1 | 36 | 32 | 32 | 88.89% | 88.89% |
| 5/19/2014 2:00 | ELA-PT-Growth an | SULLIVAN | 1 | 153 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | SULLIVAN | 1 | 166 | 150 | 141 | 90.36% | 84.94% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | SULLIVAN | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | SULLIVAN | 1 | 153 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | SULLIVAN | 1 | 166 | 152 | 148 | 91.57% | 89.16% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | SULLIVAN | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | SUMMERSVILLE R- | 1 | 27 | 19 | 19 | 70.37% | 70.37% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | SUMMERSVILLE R- | 1 | 27 | 19 | 18 | 70.37% | 66.67% |
| 5/19/2014 2:00 | ELA-PT-Technolog | TINA-AVALON R-II | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | TINA-AVALON R-II | 1 | 11 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | TINA-AVALON R-II | 1 | 12 | 12 | 12 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | TINA-AVALON R-II | 1 | 11 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | TIPTON R-VI | 1 | 42 | 42 | 39 | 100.00% | 92.86% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | TIPTON R-VI | 1 | 8 | 8 | 4 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | TIPTON R-VI | 1 | 8 | 8 | 3 | 100.00% | 37.50% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | TIPTON R-VI | 1 | 9 | 9 | 7 | 100.00% | 77.78% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | TIPTON R-VI | 1 | 8 | 8 | 2 | 100.00% | 25.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | TIPTON R-VI | 1 | 9 | 9 | 6 | 100.00% | 66.67% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | TRENTON R-IX | 1 | 77 | 63 | 63 | 81.82% | 81.82% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | TRENTON R-IX | 1 | 50 | 47 | 32 | 94.00% | 64.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | TRENTON R-IX | 1 | 27 | 25 | 25 | 92.59% | 92.59% |
| 5/19/2014 2:00 | ELA-PT-Marine An | TRI-COUNTY R-VII | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-The Ameri | TRI-COUNTY R-VII | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-The Ameri | TRI-COUNTY R-VII | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | TRI-COUNTY R-VII | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | TRI-COUNTY R-VII | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | TRI-COUNTY R-VII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | TRI-COUNTY R-VII | 1 | 5 | 2 | 2 | 40.00% | 40.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | TRI-COUNTY R-VII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-No | TRI-COUNTY R-VII | 1 | 1 | 1 | 1 | 100.00% | 100.00% |

| Date | Test | District | | Count | Col2 | Col3 | Pct1 | Pct2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G05-Math-N | TRI-COUNTY R-VII | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Commun | TROY R-III | 1 | 68 | 65 | 65 | 95.59% | 95.59% |
| 5/19/2014 2:00 | Math-PT-Making S | TROY R-III | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | TROY R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | TROY R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | TROY R-III | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | TROY R-III | 1 | 13 | 12 | 11 | 92.31% | 84.62% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | TROY R-III | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | TROY R-III | 1 | 13 | 11 | 11 | 84.62% | 84.62% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | TROY R-III | 1 | 14 | 13 | 13 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | TROY R-III | 1 | 14 | 14 | 14 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | TWIN RIVERS R-X | 1 | 46 | 46 | 46 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | TWIN RIVERS R-X | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | TWIN RIVERS R-X | 1 | 46 | 46 | 46 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | TWIN RIVERS R-X | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | TWIN RIVERS R-X | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | UNION R-XI | 1 | 214 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | UNION R-XI | 1 | 108 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | UNION R-XI | 1 | 106 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | UNION STAR R-II | 1 | 11 | 9 | 8 | 81.82% | 72.73% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | UNION STAR R-II | 1 | 11 | 10 | 10 | 90.91% | 90.91% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | UNION STAR R-II | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | UNION STAR R-II | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | UNION STAR R-II | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | UNION STAR R-II | 1 | 3 | 2 | 2 | 66.67% | 66.67% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | UNION STAR R-II | 1 | 2 | 2 | 1 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | UNION STAR R-II | 1 | 2 | 2 | 1 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | UNION STAR R-II | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | UNION STAR R-II | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | UNION STAR R-II | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Animal Jt | VALLEY PARK | 1 | 89 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | VALLEY PARK | 1 | 17 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | VALLEY PARK | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | VALLEY PARK | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | VALLEY PARK | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | VALLEY PARK | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | VALLEY R-VI | 1 | 32 | 27 | 27 | 84.38% | 84.38% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | VALLEY R-VI | 1 | 32 | 27 | 27 | 84.38% | 84.38% |
| 5/19/2014 2:00 | Math-PT-Camping | VAN BUREN R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Camping | VAN BUREN R-I | 1 | 37 | 32 | 32 | 86.49% | 86.49% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | VAN BUREN R-I | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | VAN BUREN R-I | 1 | 16 | 15 | 15 | 93.75% | 93.75% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | VAN BUREN R-I | 1 | 14 | 10 | 10 | 71.43% | 71.43% |

| Date | Assessment | District | | Count | | | Pct1 | Pct2 |
|---|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G08-Math-N | VAN BUREN R-I | 1 | 7 | 7 | 7 | 100.00% | 100.00% |
| 5/19/2014 2:00 | ELA-PT-Animals W | WARREN CO. R-III | 1 | 154 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | WARREN CO. R-III | 1 | 228 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Archeologi | WARREN CO. R-III | 1 | 25 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | WARREN CO. R-III | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | WARREN CO. R-III | 1 | 47 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Technolog | WARREN CO. R-III | 1 | 254 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | WARREN CO. R-III | 1 | 226 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | WARREN CO. R-III | 1 | 28 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | WARREN CO. R-III | 1 | 66 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Animal Ju | WARREN CO. R-III | 1 | 188 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Cell Phor | WARREN CO. R-III | 1 | 266 | 45 | 45 | 16.92% | 16.92% |
| 5/19/2014 2:00 | Math-PT-Commun | WARREN CO. R-III | 1 | 3 | 3 | 3 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Commun | WARREN CO. R-III | 1 | 99 | 85 | 85 | 85.86% | 85.86% |
| 5/19/2014 2:00 | Math-PT-Donuts | WARREN CO. R-III | 1 | 251 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | WARREN CO. R-III | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Making S | WARREN CO. R-III | 1 | 104 | 1 | 1 | 0.96% | 0.96% |
| 5/19/2014 2:00 | Math-PT-Making S | WARREN CO. R-III | 1 | 122 | 65 | 65 | 53.28% | 53.28% |
| 5/19/2014 2:00 | Math-PT-Sandbox | WARREN CO. R-III | 1 | 5 | 3 | 3 | 60.00% | 60.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | WARREN CO. R-III | 1 | 87 | 75 | 75 | 86.21% | 86.21% |
| 5/19/2014 2:00 | Math-PT-South Po | WARREN CO. R-III | 1 | 253 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | WARREN CO. R-III | 1 | 82 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-Nor | WARREN CO. R-III | 1 | 154 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WARREN CO. R-III | 1 | 24 | 23 | 23 | 95.83% | 95.83% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WARREN CO. R-III | 1 | 41 | 22 | 22 | 53.66% | 53.66% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WARREN CO. R-III | 1 | 57 | 20 | 20 | 35.09% | 35.09% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WARREN CO. R-III | 1 | 104 | 1 | 1 | 0.96% | 0.96% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | WARREN CO. R-III | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | WARREN CO. R-III | 1 | 226 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | WARREN CO. R-III | 1 | 11 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nor | WARREN CO. R-III | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WARREN CO. R-III | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WARREN CO. R-III | 1 | 66 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WARREN CO. R-III | 1 | 90 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WARREN CO. R-III | 1 | 90 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | WARREN CO. R-III | 1 | 8 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | WARREN CO. R-III | 1 | 20 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | WARREN CO. R-III | 1 | 35 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | WARREN CO. R-III | 1 | 9 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nor | WARREN CO. R-III | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARREN CO. R-III | 1 | 61 | 38 | 38 | 62.30% | 62.30% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARREN CO. R-III | 1 | 67 | 41 | 41 | 61.19% | 61.19% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARREN CO. R-III | 1 | 8 | 6 | 6 | 75.00% | 75.00% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARREN CO. R-III | 1 | 92 | 40 | 40 | 43.48% | 43.48% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARREN CO. R-III | 1 | 48 | 40 | 40 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WARREN CO. R-III | 1 | 134 | 17 | 17 | 12.69% | 12.69% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WARREN CO. R-III | 1 | 132 | 28 | 28 | 21.21% | 21.21% |
| 5/19/2014 2:00 | SBAC-G07-ELA-No | WARREN CO. R-III | 1 | 254 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | WARREN CO. R-III | 1 | 251 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | WARREN CO. R-III | 1 | 2 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | WARREN CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WARREN CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WARREN CO. R-III | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WARREN CO. R-III | 1 | 228 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WARREN CO. R-III | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WARREN CO. R-III | 1 | 18 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-Math-N | WARREN CO. R-III | 1 | 253 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Growth an | WARRENSBURG R- | 1 | 4 | 1 | 1 | 25.00% | 25.00% |
| 5/19/2014 2:00 | ELA-PT-Technology | WARRENSBURG R- | 1 | 226 | 224 | 223 | 99.12% | 98.67% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | WARRENSBURG R- | 1 | 121 | 121 | 121 | 100.00% | 100.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox | WARRENSBURG R- | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | Math-PT-Sandbox- | WARRENSBURG R- | 1 | 133 | 83 | 76 | 62.41% | 57.14% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | WARRENSBURG R- | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WARRENSBURG R- | 1 | 40 | 40 | 40 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WARRENSBURG R- | 1 | 40 | 40 | 40 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WARRENSBURG R- | 1 | 41 | 41 | 41 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARRENSBURG R- | 1 | 33 | 28 | 26 | 84.85% | 78.79% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARRENSBURG R- | 1 | 33 | 25 | 17 | 75.76% | 51.52% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARRENSBURG R- | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARRENSBURG R- | 1 | 34 | 26 | 17 | 76.47% | 50.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WARRENSBURG R- | 1 | 33 | 27 | 24 | 81.82% | 72.73% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | WARRENSBURG R- | 1 | 56 | 55 | 53 | 98.21% | 94.64% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | WARRENSBURG R- | 1 | 57 | 57 | 54 | 100.00% | 94.74% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | WARRENSBURG R- | 1 | 4 | 4 | 2 | 100.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | WARRENSBURG R- | 1 | 56 | 56 | 51 | 100.00% | 91.07% |
| 5/19/2014 2:00 | SBAC-G06-ELA-No | WARRENSBURG R- | 1 | 57 | 57 | 54 | 100.00% | 94.74% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WARRENSBURG R- | 1 | 4 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Deserts | WAYNESVILLE R-VI | 1 | 2 | 2 | 1 | 100.00% | 50.00% |
| 5/19/2014 2:00 | ELA-PT-Deserts-A | WAYNESVILLE R-VI | 1 | 287 | 240 | 226 | 83.62% | 78.75% |
| 5/19/2014 2:00 | ELA-PT-Marine An | WAYNESVILLE R-VI | 1 | 14 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Marine An | WAYNESVILLE R-VI | 1 | 14 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Space Expl | WAYNESVILLE R-VI | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Space Expl | WAYNESVILLE R-VI | 1 | 465 | 390 | 388 | 83.87% | 83.44% |
| 5/19/2014 2:00 | HS-Math-PT-Great | WAYNESVILLE R-VI | 1 | 436 | 371 | 367 | 85.09% | 84.17% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | WAYNESVILLE R-VI | 1 | 28 | 26 | 26 | 92.86% | 92.86% |
| 5/19/2014 2:00 | SBAC-G04-ELA-No | WAYNESVILLE R-VI | 1 | 58 | 50 | 48 | 86.21% | 82.76% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol | WAYNESVILLE R-VI | 1 | 57 | 52 | 51 | 91.23% | 89.47% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol | WAYNESVILLE R-VI | 1 | 2 | 2 | 2 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol | WAYNESVILLE R-VI | 1 | 58 | 47 | 45 | 81.03% | 77.59% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol | WAYNESVILLE R-VI | 1 | 57 | 53 | 51 | 92.98% | 89.47% |
| 5/19/2014 2:00 | SBAC-G04-ELA-Nol | WAYNESVILLE R-VI | 1 | 57 | 48 | 48 | 84.21% | 84.21% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WAYNESVILLE R-VI | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WAYNESVILLE R-VI | 1 | 3 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WAYNESVILLE R-VI | 1 | 14 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WAYNESVILLE R-VI | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WAYNESVILLE R-VI | 1 | 5 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WAYNESVILLE R-VI | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WAYNESVILLE R-VI | 1 | 8 | 8 | 8 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WAYNESVILLE R-VI | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WAYNESVILLE R-VI | 1 | 1 | 1 | 1 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 93 | 75 | 64 | 80.65% | 68.82% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 93 | 81 | 70 | 87.10% | 75.27% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 93 | 80 | 69 | 86.02% | 74.19% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 93 | 81 | 70 | 87.10% | 75.27% |
| 5/19/2014 2:00 | SBAC-G08-ELA-Nol | WAYNESVILLE R-VI | 1 | 93 | 77 | 65 | 82.80% | 69.89% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | WAYNESVILLE R-VI | 1 | 436 | 384 | 360 | 88.07% | 82.57% |
| 5/19/2014 2:00 | Math-PT-Animal Jt | WEBB CITY R-VII | 1 | 178 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WEBB CITY R-VII | 1 | 58 | 53 | 42 | 91.38% | 72.41% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WEBB CITY R-VII | 1 | 61 | 58 | 41 | 95.08% | 67.21% |
| 5/19/2014 2:00 | SBAC-G03-Math-N | WEBB CITY R-VII | 1 | 59 | 55 | 41 | 93.22% | 69.49% |
| 5/19/2014 2:00 | HS-ELA-PT-A New | WELLSVILLE MIDD | 1 | 23 | 22 | 22 | 95.65% | 95.65% |
| 5/19/2014 2:00 | Math-PT-Cell Phor | WELLSVILLE MIDD | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WELLSVILLE MIDD | 1 | 4 | 4 | 4 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WELLSVILLE MIDD | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WELLSVILLE MIDD | 1 | 5 | 5 | 5 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G06-Math-N | WELLSVILLE MIDD | 1 | 5 | 4 | 4 | 80.00% | 80.00% |
| 5/19/2014 2:00 | SBAC-HS-ELA-Nonl | WELLSVILLE MIDD | 1 | 23 | 22 | 14 | 95.65% | 60.87% |
| 5/19/2014 2:00 | Math-PT-Talent Sh | WENTZVILLE R-IV | 1 | 93 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WENTZVILLE R-IV | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WENTZVILLE R-IV | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WENTZVILLE R-IV | 1 | 23 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-Math-N | WENTZVILLE R-IV | 1 | 24 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-The Americ | WEST PLATTE CO. | 1 | 39 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | WEST PLATTE CO. | 1 | 57 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WEST PLATTE CO. | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WEST PLATTE CO. | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WEST PLATTE CO. | 1 | 10 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G05-ELA-Nol | WEST PLATTE CO. | 1 | 9 | 0 | 0 | 0.00% | 0.00% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5/19/2014 2:00 | SBAC-HS-Math-No | WEST PLATTE CO. | 1 | 57 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts | WEST ST. FRANCO | 1 | 74 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | Math-PT-Donuts-A | WEST ST. FRANCO | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | WEST ST. FRANCO | 1 | 74 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-G07-Math-N | WEST ST. FRANCO | 1 | 1 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | WHEATON R-III | 1 | 24 | 15 | 15 | 62.50% | 62.50% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | WHEATON R-III | 1 | 24 | 15 | 15 | 62.50% | 62.50% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | WILLARD R-II | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | ELA-PT-Uncommo | WILLARD R-II | 1 | 54 | 52 | 52 | 96.30% | 96.30% |
| 5/19/2014 2:00 | Math-PT-Turtle Ha | WILLARD R-II | 1 | 85 | 77 | 77 | 90.59% | 90.59% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WILLARD R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WILLARD R-II | 1 | 19 | 18 | 18 | 94.74% | 94.74% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WILLARD R-II | 1 | 18 | 18 | 18 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G03-ELA-No | WILLARD R-II | 1 | 2 | 1 | 1 | 50.00% | 50.00% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WILLARD R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WILLARD R-II | 1 | 17 | 14 | 14 | 82.35% | 82.35% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WILLARD R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WILLARD R-II | 1 | 17 | 16 | 16 | 94.12% | 94.12% |
| 5/19/2014 2:00 | SBAC-G04-Math-N | WILLARD R-II | 1 | 17 | 17 | 17 | 100.00% | 100.00% |
| 5/19/2014 2:00 | HS-Math-PT-Great | WINONA R-III | 1 | 38 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | SBAC-HS-Math-No | WINONA R-III | 1 | 38 | 0 | 0 | 0.00% | 0.00% |
| 5/19/2014 2:00 | ELA-PT-Renewable | WORTH CO. R-III | 1 | 32 | 31 | 31 | 96.88% | 96.88% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WORTH CO. R-III | 1 | 7 | 7 | 6 | 100.00% | 85.71% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WORTH CO. R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WORTH CO. R-III | 1 | 7 | 6 | 6 | 85.71% | 85.71% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WORTH CO. R-III | 1 | 6 | 5 | 5 | 83.33% | 83.33% |
| 5/19/2014 2:00 | SBAC-G08-ELA-No | WORTH CO. R-III | 1 | 6 | 6 | 6 | 100.00% | 100.00% |

Appendix P – 	Grade-Level Assessment Blueprints

## Blueprint for ENGLISH LANGUAGE ARTS Grades 3-4

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Reading | Literacy | 10 | 23% |
| Reading | Informational | 10 | 23% |
| Writing | Organization/Purpose | 0-5 | 0-11% |
| Writing | Evidence/Elaboration | 0-5 | 0-11% |
| Writing | Conventions | 5 | 11% |
| Speaking/Listening | Listening | 9 | 20% |
| Research | Research | 5 | 11% |
| Total | | 44 | 100% |

## Blueprint for ENGLISH LANGUAGE ARTS Grade 5

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Reading | Literacy | 10 | 20% |
| Reading | Informational | 10 | 20% |
| Writing | Organization/Purpose | 1-6 | 2-12% |
| Writing | Evidence/Elaboration | 1-6 | 2-12% |
| Writing | Conventions | 6 | 12% |
| Speaking/Listening | Listening | 9 | 18% |
| Research | Research | 8 | 16% |
| Total | | 50 | 100% |

## Blueprint for ENGLISH LANGUAGE ARTS Grades 6-7

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Reading | Literacy | 9 | 20% |
| Reading | Informational | 12 | 27% |
| Writing | Organization/Purpose | 0-5 | 0-11% |
| Writing | Evidence/Elaboration | 0-5 | 0-11% |
| Writing | Conventions | 5 | 11% |
| Speaking/Listening | Listening | 9 | 20% |
| Research | Research | 5 | 11% |
| Total | | 45 | 100% |

## Blueprint for ENGLISH LANGUAGE ARTS Grade 8

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Reading | Literacy | 9 | 18% |
| Reading | Informational | 12 | 24% |
| Writing | Organization/Purpose | 1-6 | 2-12% |
| Writing | Evidence/Elaboration | 1-6 | 2-12% |
| Writing | Conventions | 6 | 12% |
| Speaking/Listening | Listening | 9 | 18% |
| Research | Research | 8 | 16% |
| **Total** | | 51 | 100% |

## Blueprint for MATHEMATICS Grades 3-4

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Concepts And Procedures | Priority Cluster | 15 | 48% |
| Concepts And Procedures | Supporting Cluster | 5 | 16% |
| Problem Solving | Problem Solving | 0-5 | 0-16% |
| Communicating Reasoning | Communicating Reasoning | 6 | 19% |
| Modeling And Data Analysis | Modeling And Data Analysis | 0-5 | 0-16% |
| **Total** | | 31 | 100% |

## Blueprint for MATHEMATICS Grade 5

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Concepts And Procedures | Priority Cluster | 15 | 41% |
| Concepts And Procedures | Supporting Cluster | 5 | 14% |
| Problem Solving | Problem Solving | 0-9 | 0-29% |
| Communicating Reasoning | Communicating Reasoning | 8 | 26% |
| Modeling And Data Analysis | Modeling And Data Analysis | 0-9 | 0-29% |
| **Total** | | 37 | 100% |

## Blueprint for MATHEMATICS Grades 6-7

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Concepts And Procedures | Priority Cluster | 14-15 | 45-48% |
| Concepts And Procedures | Supporting Cluster | 5 | 16% |
| Problem Solving | Problem Solving | 0-5 | 0-16% |
| Communicating Reasoning | Communicating Reasoning | 6 | 19% |
| Modeling And Data Analysis | Modeling And Data Analysis | 0-5 | 0-16% |
| **Total** | | 30-31 | 100% |

## Blueprint for MATHEMATICS Grade 8

| Claim | Category | Point Range | Range Of Emphasis |
|---|---|---|---|
| Concepts And Procedures | Priority Cluster | 14-15 | 38-41% |
| Concepts And Procedures | Supporting Cluster | 5 | 14% |
| Problem Solving | Problem Solving | 0-9 | 0-24% |
| Communicating Reasoning | Communicating Reasoning | 8 | 22% |
| Modeling And Data Analysis | Modeling And Data Analysis | 0-9 | 0-24% |
| Total | | 36-37 | 100% |

## Blueprint for SCIENCE Grade 5

| Content Strand | Point Range | Range Of Emphasis |
|---|---|---|
| 1. ME: Properties And Principles Of Matter And Energy | 7-8 | 12-13% |
| 2. FM: Properties And Principles Of Force And Motion | 5-6 | 8-10% |
| 3. LO: Characteristics and Interactions of Living Organisms | 5-6 | 8-10% |
| 4. EC: Changes in Ecosystems and Interactions of Organisms with their Environments | 5-7 | 8-12% |
| 5. ES: Processes And Interactions Of The Earth's Systems | 7-10 | 12-17% |
| 6. UN: Composition And Structure Of The Universe And The Motions Of The Objects Within It | 6-7 | 10-12% |
| 7. IN: Processes Of Scientific Inquiry | 14-17 | 23-28% |
| 8. ST: Impact Of Science, Technology, And Human Activity | 5-6 | 8-10% |
| Total | 60 | 100% |

## Blueprint for SCIENCE Grade 8

| Content Strand | Point Range | Range Of Emphasis |
|---|---|---|
| 1. ME: Properties And Principles Of Matter And Energy | 7-8 | 12-13 |
| 2. FM: Properties And Principles Of Force And Motion | 5-6 | 8-10% |
| 3. LO: Characteristics and Interactions of Living Organisms | 5-6 | 8-10% |
| 4. EC: Changes in Ecosystems and Interactions of Organisms with their Environments | 5-7 | 8-12% |
| 5. ES: Processes And Interactions Of The Earth's Systems | 7-10 | 12-17% |
| 6. UN: Composition And Structure Of The Universe And The Motions Of The Objects Within It | 6-7 | 10-12% |
| 7. IN: Processes Of Scientific Inquiry | 14-17 | 23-28% |
| 8. ST: Impact Of Science, Technology, And Human Activity | 5-6 | 8-10% |
| Total | 60 | 100% |

Appendix Q – 2014 MAP-A Field Test School Districts and Charter Schools

District
Academie Lafayette
Advance R-IV
Affton 101
Arcadia Valley R-II
Aurora R-VIII
AVA R-I
Bayless
Belton 124
Bevier C-4
Bismarck R-V
Blue Eye R-V
Blue Springs R-IV
Boonville R-I
BOWLING GREEN R-I
Branson R-IV
Braymer C-4
Buchanan Co. R-IV
Bunker R-III
Callao C-8
Camdenton R-III
Campbell R-II
Carl Junction R-I
Carrollton R-VII
Carthage R-IX
Center 58
Chaffee R-II
Chilhowee R-IV
Chillicothe R-II
Clark Co. R-I
Clayton
Clearwater R-I
Clinton
Clinton Co. R-III
COLE CAMP R-I
Columbia 93
Concordia R-II
Confluence Academies
Crawford Co. R-I
Dallas Co. R-I
Dent-Phelps R-III
Dexter R-Xi
Dixon R-1
Dunklin R-V
East Newton Co. R-VI
East Prairie R-2
Elsberry R-II

Ewing Marion Kauffman School
FAIRVIEW R-XI
Farmington R-VII
Ferguson-Florissant R-II
Forsyth R-III
Fox C-6
Francis Howell R-III
Ft. Zumwalt R-II
Gainesville R-V
Galena R-II
Gallatin R-V
Gasconade Co. R-I
Gasconade Co. R-II
Gilman City R-IV
Glasgow
Grain Valley R-V
Greenfield R-IV
Greenville R-II
Hale R-I
Hamilton R-II
Hancock Place
Harrisonville R-IX
Hazelwood
Henry Co. R-I
Hickory Co. R-I
Independence 30
Jasper Co. R-V
Jennings
Johnson Co. R-VII
Joplin Schools
Kearney R-I
Kennett 39
Kingston K-14
KIRKSVILLE R-III
Kirkwood R-VII
Knob Noster R-VIII
Knox Co. R-I
La Monte R-IV
La Plata R-II
Laclede Co. R-I
Ladue
Lafayette Co. C-1
LAMAR R-I
Lawson R-Xiv
Lebanon R-III
Lee's Summit R-VII
Lewis Co. C-1

Licking R-VIII
LINCOLN R-II
Lindbergh Schools
Logan-Rogersville R-VIII
Lone Jack C-6
Lonedell R-Xiv
Louisiana R-II
Macon Co. R-I
Madison C-3
Malden R-I
Malta Bend R-V
Maplewood-Richmond Heights
Marceline R-V
Maries Co. R-II
Marshfield R-I
Maryville R-II
Mcdonald Co. R-I
MEADOW HEIGHTS R-II
Mehlville R-IX
Miller R-II
Missouri
Mo Schls For The Sev Disabled
Mo School For The Blind
Monett R-I
Monroe City R-I
Montgomery Co. R-II
Morgan Co. R-II
Mound City R-II
Neosho R-V
Nevada R-V
New Franklin R-I
New Haven
Niangua R-V
North Callaway Co. R-I
North Mercer Co. R-III
North Nodaway Co. R-VI
North Platte Co. R-I
North St. Francois Co. R-I
Northeast Vernon Co. R-I
Northwest R-I
Norwood R-I
Oak Grove R-VI
Odessa R7 Schools
Orchard Farm School District
Oregon-Howell R-III
Otterville R-VI
Ozark R-VI

Palmyra R-I
Parkway C-2
Pattonville R-III
Pemiscot Co. Spec. Sch. Dist.
Perry Co. 32
Phelps Co. R-III
Pierce City R-VI
Pike Co. R-III
Pleasant Hill R-III
Poplar Bluff R-I
Putnam Co. R-I
Puxico R-VIII
Ralls Co. R-II
Raymore-Peculiar R-II
Raytown C-2
Republic R-III
Richards R-V
Richland R-I
Richmond R-Xvi
Ritenour
Rockwood R-VI
Rolla 31
Salem R-80
Sarcoxie R-II
School Of The Osage
Schuyler Co. R-I
Scotland Co. R-I
Scott City R-I
Scott Co. Central
Scott Co. R-IV
Sedalia 200
Seymour R-II
Shelby Co. R-IV
Sheldon R-VIII
Shell Knob 78
Sherwood Cass R-VIII
Sikeston R-6
Silex R-I
SKYLINE R-II
Smithton R-VI
South Callaway Co. R-II
South Harrison Co. R-II
South Holt Co. R-I
SOUTH NODAWAY CO. R-IV
Southern Boone Co. R-I
Southland C-9
Southwest R-V

Specl. Sch. Dst. St. Louis Co.
Springfield R-XII
St. Charles R-VI
St. Clair R-XIII
St. James R-I
St. Louis City
Stockton R-I
Strafford R-VI
Sullivan
Sunrise R-IX
Tipton R-VI
Trenton R-IX
Twin Rivers R-X
Union R-Xi
Union Star R-II
University City
Valley Park
VAN-FAR R-I
Verona R-VII
Warren Co. R-III
Warrensburg R-VI
Waynesville R-VI
Webster Groves
West Plains R-VII
Willard R-II
Willow Springs R-IV
WINDSOR C-1
Winfield R-IV
Winona R-III
Wright City R-II
Zalma R-V

Appendix R –     MAP-A ELA and MA Test Blueprints

# DLM Mathematics Integrated Assessment Model
## 2014-15 Blueprint

In this document, the "blueprint" refers to the pool of available Essential Elements (EEs) and the requirements for coverage within each conceptual area. A general description of the content covered is provided for each grade. The specific options and minimum expectations for each student's assessment are provided with each table. Educators should consult their state department of education for additional guidance on selecting content.

The specific EEs available in each grade are listed in tables beginning on the next page. EEs are organized according to conceptual area.

## Major Claims and Conceptual Areas in Mathematics

| Major Claim | Conceptual Area | |
|---|---|---|
| 1. Students demonstrate increasingly complex understanding of number sense. | M.C1.1 | Understand number structures (counting, place value, fraction) |
| | M.C1.2 | Compare, compose, and decompose numbers and sets |
| | M.C1.3 | Calculate accurately and efficiently using simple arithmetic operations |
| 2. Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles. | M.C2.1 | Understand and use geometric properties of two- and three-dimensional shapes |
| | M.C2.2 | Solve problems involving area, perimeter, and volume |
| 3. Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures. | M.C3.1 | Understand and use measurement principles and units of measure |
| | M.C3.2 | Represent and interpret data displays |
| 4. Students solve increasingly complex mathematical problems, making productive use of algebra and functions. | M.C4.1 | Use operations and models to solve problems |
| | M.C4.2 | Understand patterns and functional thinking |

## Grade 3: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | **Students demonstrate increasingly complex understanding of number sense.** |
| | **Choose two EEs from Claim 1 in different conceptual areas, i.e., one EE in C1.1 and one EE in C1.3.** |
| | **M.C1.1** | 3.NBT.2 | Demonstrate understanding of place value to tens. |
| | | 3.NBT.3 | Count by tens using models such as objects, base ten blocks, or money. |
| | | 3.NF.1-3 | Differentiate a fractional part from a whole. |
| | **M.C1.3** | 3.OA.4 | Solve addition and subtraction problems when result is unknown, limited to operands and results within 20. |
| **2** | **Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.** |
| | **All students are assessed on the EE in Claim 2.** |
| | **M.C2.2** | 3.G.2 | Recognize that shapes can be partitioned into equal areas. |
| **3** | **Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures.** |
| | **Choose two EEs from Claim 3.** |
| | **M.C3.1** | 3.MD.1 | Tell time to the hour on a digital clock. |
| | | 3.MD.4 | Measure length of objects using standard tools, such as rulers, yardsticks, and meter sticks |
| | **M.C3.2** | 3.MD.3 | Use picture or bar graph data to answer questions about data. |
| **4** | **Students solve increasingly complex mathematical problems, making productive use of algebra and functions.** |
| | **Choose one EE from Claim 4.** |
| | **M.C4.1** | 3.OA.1-2 | Use repeated addition to find the total number of objects and determine the sum. |
| | | 3.OA.8 | Solve one-step real world problems using addition or subtraction within 20. |
| | **M.C4.2** | 3.OA.9 | Identify arithmetic patterns. |

## Grade 4: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | Students demonstrate increasingly complex understanding of number sense. Choose two EEs from Claim 1 in different conceptual areas. | | |
| | **M.C1.1** | 4.NF.1-2 | Identify models of one half (1/2) and one fourth (1/4). |
| | | 4.NF.3 | Differentiate between whole and half. |
| | **M.C1.2** | 4.NBT.2 | Compare whole numbers to 10 using symbols (<, >, =). |
| | | 4.NBT.3 | Round any whole number 0-30 to the nearest ten. |
| | **M.C1.3** | 4.NBT.4 | Add and subtract two-digit whole numbers. |
| **2** | Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles. Choose two EEs from Claim 2 in different conceptual areas. | | |
| | **M.C2.1** | 4.G.1 | Recognize parallel lines and intersecting lines. |
| | | 4.MD.5 | Recognize angles in geometric shapes. |
| | | 4.MD.6 | Identify angles as larger and smaller. |
| | **M.C2.2** | 4.MD.3 | Determine the area of a square or rectangle by counting units of measure (unit squares). |
| **3** | Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures. Choose two EEs from Claim 3 in different conceptual areas. | | |
| | **M.C3.1** | 4.MD.2.a | Tell time using a digital clock. Tell time to the nearest hour using an analog clock. |
| | | 4.MD.2.b | Measure mass or volume using standard tools. |
| | | 4.MD.2.d | Identify coins (penny, nickel, dime, quarter) and their values. |
| | **M.C3.2** | 4.MD.4.b | Interpret data from a picture or bar graph. |
| **4** | Students solve increasingly complex mathematical problems, making productive use of algebra and functions. Choose two EEs from Claim 4 in different conceptual areas. | | |
| | **M.C4.1** | 4.OA.1-2 | Demonstrate the connection between repeated addition and multiplication. |
| | | 4.OA.3 | Solve one-step real-world problems using addition or subtraction within 100. |
| | **M.C4.2** | 4.OA.5 | Use repeating patterns to make predictions. |

## Grade 5: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | | | **Students demonstrate increasingly complex understanding of number sense.**<br>**Choose three EEs from Claim 1 in at least two different conceptual areas.** |
| | **M.C1.1** | 5.NF.1 | Identify models of halves (1/2, 2/2) and fourths (1/4, 2/4, 3/4, 4/4). |
| | | 5.NF.2 | Identify models of thirds (1/3, 2/3, 3/3) and tenths (1/10, 2/10, 3/10, 4/10, 5/10, 6/10, 7/10, 8/10, 9/10, 10/10). |
| | **M.C1.2** | 5.NBT.1 | Compare numbers up to 99 using base ten models. |
| | | 5.NBT.3 | Compare whole numbers up to 100 using symbols (<, >, =). |
| | | 5.NBT.4 | Round two-digit whole numbers to the nearest 10 from 0—90. |
| | **M.C1.3** | 5.NBT.5 | Multiply whole numbers up to 5x5. |
| | | 5.NBT.6-7 | Illustrate the concept of division using fair and equal shares. |
| **2** | | | **Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.**<br>**Choose one EE from Claim 2.** |
| | **M.C2.1** | 5.G.1-4 | Sort two-dimensional figures and identify the attributes (angles, number of sides, corners, color) they have in common. |
| | | 5.MD.3 | Identify common three-dimensional shapes. |
| | **M.C2.2** | 5.MD.4-5 | Determine the volume of a rectangular prism by counting units of measure (unit cubes). |
| **3** | | | **Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures.**<br>**Choose two EEs from Claim 3 in different conceptual areas.** |
| | **M.C3.1** | 5.MD.1.a | Tell time using an analog or digital clock to the half or quarter hour. |
| | | 5.MD.1.b | Use standard units to measure weight and length of objects. |
| | | 5.MD.1.c | Indicate relative value of collections of coins. |
| | **M.C3.2** | 5.MD.2 | Represent and interpret data on a picture, line plot, or bar graph. |
| **4** | | | **Students solve increasingly complex mathematical problems, making productive use of algebra and functions.**<br>**All students are assessed on the EE in C4.2.** |
| | **M.C4.2** | 5.OA.3 | Identify and extend numerical patterns. |

## Grade 6: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | Students demonstrate increasingly complex understanding of number sense.<br>Choose two EEs from Claim 1 in different conceptual areas. | | |
| | **M.C1.1** | 6.RP.1 | Demonstrate a simple ratio relationship. |
| | **M.C1.2** | 6.NS.1 | Compare the relationships between two unit fractions. |
| | | 6.NS.5-8 | Understand that positive and negative numbers are used together to describe quantities having opposite directions or values (e.g., temperature above/below zero). |
| | **M.C1.3** | 6.NS.2 | Apply the concept of fair share and equal shares to divide. |
| | | 6.NS.3 | Solve two-factor multiplication problems with products up to 50 using concrete objects and/or a calculator. |
| **2** | Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.<br>Choose one EE from Claim 2. | | |
| | **M.C2.2** | 6.G.1 | Solve real-world and mathematical problems about area using unit squares. |
| | | 6.G.2 | Solve real-world and mathematical problems about volume using unit cubes. |
| **3** | Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures.<br>Choose one EE from Claim 3. | | |
| | **M.C3.2** | 6.SP.5 | Summarize data distributions shown in graphs or tables. |
| **4** | Students solve increasingly complex mathematical problems, making productive use of algebra and functions.<br>Choose two EEs from Claim 4. | | |
| | **M.C4.1** | 6.EE.1-2 | Identify equivalent number sentences. |
| | | 6.EE.3 | Apply the properties of addition to identify equivalent numerical expressions. |
| | | 6.EE.5-7 | Match an equation to a real-world problem in which variables are used to represent numbers. |

## Grade 7: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | Students demonstrate increasingly complex understanding of number sense. Choose three EEs in Claim 1; at least one in C1.1 and at least one in C1.3. | | |
| | **M.C1.1** | 7.NS.2.c-d | Express a fraction with a denominator of 10 as a decimal. |
| | | 7.RP.1-3 | Use a ratio to model or describe a relationship. |
| | **M.C1.2** | 7.NS.3 | Compare quantities represented as decimals in real world examples to tenths. |
| | **M.C1.3** | 7.NS.1 | Add fractions with like denominators (halves, thirds, fourths, and tenths) with sums less than or equal to one. |
| | | 7.NS.2.a | Solve multiplication problems with products to 100 |
| | | 7.NS.2.b | Solve division problems with divisors up to five and also with a divisor of 10 without remainders |
| **2** | Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles. Choose two EEs in Claim 2 in different conceptual areas. | | |
| | **M.C2.1** | 7.G.1 | Match two similar geometric shapes that are proportional in size and in the same orientation. |
| | | 7.G.2 | Recognize geometric shapes with given conditions. |
| | | 7.G.5 | Recognize angles that are acute, obtuse, and right. |
| | **M.C2.2** | 7.G.4 | Determine the perimeter of a rectangle by adding the measures of the sides. |
| **3** | Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures. Choose one EE from Claim 3. | | |
| | **M.C3.2** | 7.SP.3 | Compare two sets of data within a single data display such as a picture graph, line plot, or bar graph. |
| | | 7.SP.5-7 | Describe the probability of events occurring as possible or impossible. |
| **4** | Students solve increasingly complex mathematical problems, making productive use of algebra and functions. Choose one EE from Claim 4. | | |
| | **M.C4.1** | 7.EE.1 | Use the properties of operations as strategies to demonstrate that expressions are equivalent. |
| | **M.C4.2** | 7.EE.2 | Identify an arithmetic sequence of whole numbers with a whole number common difference. |

## Grade 7: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description |
|---|---|---|---|
| **1** | Students demonstrate increasingly complex understanding of number sense. Choose two EEs in Claim 1 in different conceptual areas. | | |
| | **M.C1.1** | 8.NS.2.a | Express a fraction with a denominator of 100 as a decimal. |
| | **M.C1.2** | 8.NS.2.b | Compare quantities represented as decimals in real-world examples to hundredths. |
| | **M.C1.3** | 8.EE.1 | Identify the meaning of an exponent (limited to exponents of 2 and 3). |
| | | 8.NS.1 | Subtract fractions with like denominators (halves, thirds, fourths, and tenths) with minuends less than or equal to one. |
| **2** | Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles. Choose two EEs in Claim 2 in different conceptual areas. | | |
| | **M.C2.1** | 8.G.1 | Recognize translations, rotations, and reflections of shapes. |
| | | 8.G.2 | Identify shapes that are congruent. |
| | | 8.G.4 | Identify similar shapes with and without rotation. |
| | | 8.G.5 | Compare any angle to a right angle and describe the angle as greater than, less than, or congruent to a right angle. |
| | **M.C2.2** | 8.G.9 | Use the formulas for perimeter, area, and volume to solve real-world and mathematical problems (limited to perimeter and area of rectangles and volume of rectangular prisms). |
| **3** | Students demonstrate Increasingly complex understanding of measurement, data, and analytic procedures. All students are assessed on the EE from C3.2. | | |
| | **M.C3.2** | 8.SP.4 | Construct a graph or table from given categorical data and compare data categorized in the graph or table. |
| **4** | Students solve increasingly complex mathematical problems, making productive use of algebra and functions. Choose two EEs from Claim 4. | | |
| | **M.C4.1** | 8.EE.7 | Solve simple algebraic equations with one variable using addition and subtraction. |
| | **M.C4.2** | 8.EE.2 | Identify a geometric sequence of whole numbers with a whole number common ratio. |
| | | 8.F.1-3 | Given a function table containing at least 2 complete ordered pairs, identify a missing number that completes another ordered pair (limited to linear functions). |
| | | 8.F.4 | Determine the values or rule of a function using a graph or a table. |

# High School: Available Essential Elements and minimum expectation for each student's assessment

| Claim | Conceptual Area | EE | Description | Available Math 9 | Available Math 10 | Available Math 11 and 12 |
|---|---|---|---|---|---|---|
| — **Choose a minimum of six EEs across a minimum of three Claims *(see next page for Claim 4)*.** | | | | | | |
| 1 | M.C1.3 | N-CN.2.a | Use the commutative, associative, and distributive properties to add, subtract, and multiply whole numbers. | ● | | |
| | | N-CN.2.b | Solve real-world problems involving addition and subtraction of decimals, using models when needed. | ● | | |
| | | N-CN.2.c | Solve real-world problems involving multiplication of decimals and whole numbers, using models when needed. | ● | | |
| | | N-RN.1 | Determine the value of a quantity that is squared or cubed. | | | ● |
| | | S-CP.1-5 | Identify when events are independent or dependent. | | ● | |
| | | S-IC.1-2 | Determine the likelihood of an event occurring when the outcomes are equally likely to occur. | | | ● |
| 2 | M.C2.1 | G-CO.1 | Know the attributes of perpendicular lines, parallel lines, and line segments; angles, and circles. | ● | | |
| | | G-CO.4-5 | Given a geometric figure and a rotation, reflection, or translation of that figure, identify the components of the two figures that are congruent. | | ● | |
| | | G-CO.6-8 | Identify corresponding congruent and similar parts of shapes. | | | ● |
| | | G-MG.1-3 | Use properties of geometric shapes to describe real-life objects. | ● | | |
| | M.C2.2 | G-GPE.7 | Find perimeter and area of squares and rectangles to solve real-world problems. | ● | | |
| 3 | M.C3.1 | N-Q.1-3 | Express quantities to the appropriate precision of measurement. | | ● | |
| | M.C3.2 | S-ID.1-2 | Given data, construct a simple graph (table, line, pie, bar, or picture) and interpret the data. | | ● | |
| | | S-ID.3 | Interpret general trends on a graph or chart. | | | ● |
| | | S-ID.4 | Calculate the mean of a given data set (limit the number of data points to fewer than five). | | ● | |

| Claim | Conceptual Area | EE | Description | Available Math 9 | Available Math 10 | Available Math 11 and 12 |
|---|---|---|---|---|---|---|
| **4** | **M.C4.1** | A-CED.1 | Create an equation involving one operation with one variable, and use it to solve a real-world problem. | | ● | |
| | | A-CED.2-4 | Solve one-step inequalities. | | ● | |
| | | A-SSE.1 | Identify an algebraic expression involving one arithmetic operation to represent a real-world problem. | ● | | |
| | | A-SSE.3 | Solve simple algebraic equations with one variable using multiplication and division. | ● | | |
| | **M.C4.2** | A-REI.10-12 | Interpret the meaning of a point on the graph of a line. | | ● | |
| | | A-SSE.4 | Determine the successive term in a geometric sequence given the common ratio. | | | ● |
| | | F-BF.1 | Select the appropriate graphical representation (first quadrant) given a situation involving constant rate of change. | | ● | |
| | | F-BF.2 | Determine an arithmetic sequence with whole numbers when provided a recursive rule. | | | ● |
| | | F-IF.1-3 | Use the concept of function to solve problems. | | | ● |
| | | F-IF.4-6 | Construct graphs that represent linear functions with different rates of change and interpret which is faster/slower, higher/lower, etc. | | | ● |
| | | F-LE.1-3 | Model a simple linear function such as y=mx to show that these functions increase by equal amounts over equal intervals. | | | ● |

# DLM English Language Arts Integrated Assessment Model
# 2014-15 Blueprint

In this document, the "blueprint" refers to the pool of available Essential Elements (EEs) and the requirements for coverage within each conceptual area. A general description of the content covered is provided for each grade. The specific options and minimum expectations for each student's assessment are provided with each table. Educators should consult their state department of education for additional guidance on selecting content.

The specific EEs available in each grade are listed in tables beginning on the next page. EEs are organized according to conceptual area.

## Major Claims and Conceptual Areas in ELA

| Major Claim | Conceptual Area | |
|---|---|---|
| Students can comprehend text in increasingly complex ways | ELA.C1.1 | Determine critical elements of text |
| | ELA.C1.2 | Construct understandings of text |
| | ELA.C1.3 | Integrate ideas and information from text |
| Students can produce writing for a range of purposes and audiences | ELA.C2.1 | Use writing to communicate |
| | ELA.C2.2 | Integrate  ideas and Information in writing |
| Students can communicate for a range of purposes and audiences | ELA.C3.1 | Use language to communicate with others |
| | ELA.C3.2 | Clarify and contribute in discussion |
| Students can investigate topics and present information | ELA.C4.1 | Use sources and information |
| | ELA.C4.2 | Collaborate and present ideas |

## Grade 3: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | DESCRIPTION |
|---|---|---|
| **ELA.C1.1** | **Choose at least three EEs, including at least one RL and one RI.** | |
| | EE.RL.3.1 | Answer who and what questions to demonstrate understanding of details in a text. |
| | EE.RL.3.2 | Associate details with events in stories from diverse cultures. |
| | EE.RL.3.3 | Identify the feelings of characters in a story. |
| | EE.RL.3.5 | Determine the beginning, middle, and end of a familiar story with a logical order. |
| | EE.RI.3.1 | Answer who and what questions to demonstrate understanding of details in a text. |
| | EE.RI.3.2 | Identify details in a text. |
| | EE.RI.3.3 | Order two events from a text as "first" and "next". |
| | EE.RI.3.5 | With guidance and support, use text features including headings and key words to locate information in a text. |
| **ELA.C1.2** | **Choose two EEs in C1.2 (L, RL or RI) – EEs must be from different strands, i.e. RL and L, not RL and RL.** | |
| | EE.RL.3.4 | Determine words and phrases that complete literal sentences in a text. |
| | EE.RI.3.4 | Determine words and phrases that complete literal sentences in a text. |
| | EE.RI.3.8 | Identify two related points the author makes in an informational text. |
| | EE.L.3.5.a | Determine the literal meaning of words and phrases in context. |
| | EE.L.3.5.c | Identify words that describe personal emotional states. |
| **ELA.C1.3** | **Choose at least one EE (RL or RI).** | |
| | EE.RL.3.9 | Identify common elements in two stories in a series. |
| | EE.RI.3.9 | Identify similarities between two texts on the same topic. |
| **ELA.C2.1** | **All students are assessed in both of these EEs through the writing assessment.** | |
| | EE.W.3.2.a | Select a topic and write about it including one fact or detail. |
| | EE.W.3.4 | With guidance and support produce writing that expresses more than one idea. |

# Grade 4: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.1** | **Choose at least three EEs in C1.1, including at least one RL and one RI.** | |
| | EE.RL.4.1 | Use details from the text to recount what the text says. |
| | EE.RL.4.3 | Use details from the text to describe characters in the story. |
| | EE.RL.4.5 | Identify elements that are characteristic of stories. |
| | EE.RI.4.1 | Identify explicit details in an informational text. |
| | EE.RI.4.2 | Identify the main idea of a text when it is explicitly stated. |
| | EE.RI.4.3 | Identify an explicit detail that is related to an individual, event or idea in a historical, scientific, or technical text. |
| | EE.RI.4.5 | Identify elements that are characteristic of informational texts. |
| **ELA.C1.2** | **Choose two EEs in C1.2 (L, RL or RI) – EEs must be from different strands, i.e. RL and L, not RL and RL.** | |
| | EE.RL.4.2 | Identify the theme or central idea of a familiar story, drama or poem. |
| | EE.RL.4.4 | Determine the meaning of words in a text. |
| | EE.RL.4.6 | Identify the narrator of a story. |
| | EE.RI.4.4 | Determine meaning of words in text. |
| | EE.RI.4.8 | Identify one or more reasons supporting a specific point in an informational text. |
| | EE.L.4.5.c | Demonstrate an understanding of opposites. |
| **ELA.C1.3** | **All students are assessed in this EE for C1.3** | |
| | EE.RI.4.9 | Compare details presented in two texts on the same topic. |
| **ELA.2.1** | **All students are assessed in these three EEs as part of the writing assessment.** | |
| | EE.L.4.2.a | Capitalize the first word in a sentence. |
| | EE.L.4.2.d | Spell words phonetically, drawing on knowledge of letter-sound relationships, and/or common spelling patterns |
| | EE.W.4.2.b | List words, facts, or details related to the topic. |

## Grade 5: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.1** | **Choose at least two EEs in C1.1, including at least one RL and one RI** | |
| | EE.RL.5.1 | Identify words in the text to answer a question about explicit information. |
| | EE.RI.5.1 | Identify words in the text to answer a question about explicit information. |
| | EE.RI.5.5 | Determine if a text tells about events, gives directions, or provides information on a topic. |
| | EE.RI.5.7 | Locate information in print or digital sources. |
| **ELA.C1.2** | **Choose three EE's in C1.2 (L, RL, or RI) – EEs must be from at least two different strands** | |
| | EE.RL.5.2 | Identify the central idea or theme of a story, drama or poem. |
| | EE.RL.5.4 | Determine the intended meaning of multi-meaning words in a text. |
| | EE.RL.5.6 | Determine the point of view of the narrator. |
| | EE.RI.5.2 | Identify the main idea of a text when it is not explicitly stated. |
| | EE.RI.5.4 | Determine the meanings of domain-specific words and phrases. |
| | EE.RI.5.8 | Identify the relationship between a specific point and supporting reasons in an informational text. |
| | EE.L.5.4.a | Use sentence level context to determine which word is missing from a content area text. |
| | EE.L.5.5.c | Demonstrate understanding of words that have similar meanings. |
| **ELA.C1.3** | **Choose at least one EE in C1.3 (RL or RI)** | |
| | EE.RL.5.3 | Compare two characters in a familiar story. |
| | EE.RL.5.5 | Identify story element that undergoes change from beginning to end. |
| | EE.RL.5.9 | Compare stories, myths, or texts with similar topics or themes. |
| | EE.RI.5.3 | Compare two individuals, events or ideas in a text. |
| | EE.RI.5.9 | Compare and contrast details gained from two texts on the same topic. |
| **ELA.C2.1** | **All students are assessed in both of these EEs through the writing assessment.** | |
| | EE.W.5.2.b | Provide facts, details, or other information related to the topic. |
| | EE.W.5.2.a | Introduce a topic and write to convey information about it including visual, tactual, or multimedia information as appropriate. |

## Grade 6: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.1** | **All students are assessed in this EE for C1.1** | |
| | EE.RI.6.5 | Determine how the title fits the structure of the text. |
| **ELA.C1.2** | **Choose three EEs in C1.2 (L, RL or RI) –EEs must be from at least two different strands.** | |
| | EE.RL.6.1 | Determine what a text says explicitly as well as what simple inferences must be drawn. |
| | EE.RL.6.2 | Identify details in a text that are related to the theme or central idea. |
| | EE.RL.6.4 | Determine how word choice changes the meaning in a text. |
| | EE.RL.6.6 | Identify words or phrases in the text that describe or show what the narrator or speaker is thinking or feeling. |
| | EE.RI.6.1 | Analyze a text to determine what it says explicitly as well as what inferences should be drawn. |
| | EE.RI.6.2 | Determine the main idea of a passage and details or facts related to it. |
| | EE.RI.6.4 | Determine how word choice changes the meaning of a text. |
| | EE.RI.6.6 | Identify words or phrases in the text that describe or show the author's point of view. |
| | EE.RI.6.8 | Distinguish claims in a text supported by reason. |
| | EE.L.6.5.a | Identify the meaning of simple similes (e.g., The man was as big as a tree.). |
| | EE.L.6.5.b | Demonstrate understanding of words by identifying other words with similar and different meanings. |
| **ELA.C1.3** | **Choose at least two EEs in C1.3, including at least one RL and one RI.** | |
| | EE.RL.6.3 | Can identify how a character responds to a challenge in story. |
| | EE.RL.6.5 | Determine the structure of a text (e.g., story, poem, or drama). |
| | EE.RI.6.3 | Identify a detail that elaborates upon individuals, events, or ideas introduced in a text. |
| | EE.RI.6.9 | Compare and contrast how two texts describe the same event. |
| **ELA.C2.1** | **All students are assessed all three of these EEs through the writing assessment.** | |
| | EE.L.6.2.b | Spell untaught words phonetically, drawing on letter-sound relationships and common spelling patterns. |
| | EE.W.6.2.a | Introduce a topic and write to convey ideas and information about it including visual, tactual, or multimedia information as appropriate. |
| | EE.W.6.2.b | Provide facts, details, or other information related to the topic. |

## Grade 7: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.1** | **All students are assessed in this EE for C1.1** | |
| | EE.RI.7.5 | Determine how a fact, step, or event fits into the overall structure of the text. |
| **ELA.C1.2** | **Choose at least three EEs in C1.2 (at least one RL and one RI)** | |
| | EE.RL.7.1 | Analyze text to identify where information is explicitly stated and where inferences must be drawn. |
| | EE.RL.7.2 | Identify events in a text that are related to the theme or central idea. |
| | EE.RL.7.4 | Determine the meaning of simple idioms and figures of speech as they are used in a text. |
| | EE.RI.7.1 | Analyze text to identify where information is explicitly stated and where inferences must be drawn. |
| | EE.RI.7.2 | Determine two or more central ideas in a text. |
| | EE.RI.7.4 | Determine how words or phrases are used to persuade or inform a text. |
| | EE.RI.7.6 | Determine an author's purpose or point of view. |
| | EE.RI.7.8 | Determine how a claim or reason fits into the overall structure of an informational text. |
| **ELA.C1.3** | **Choose at least two EEs in C1.3, including at least one RL and one RI.** | |
| | EE.RL.7.3 | Determine how two or more story elements are related. |
| | EE.RL.7.5 | Compare the structure of two or more texts (e.g., stories, poems, or dramas). |
| | EE.RI.7.3 | Determine how two individuals, events or ideas in a text are related. |
| | EE.RI.7.9 | Compare and contrast how different texts on the same topic present the details. |
| **ELA.C2.1** | **All students are assessed in five of these EEs through the writing assessment.** | |
| | EE.L.7.2.a | Use end punctuation when writing a sentence or question. |
| | EE.L.7.2.b | Spell words phonetically, drawing on knowledge of letter-sound relationships and/or common spelling patterns. |
| | EE.W.7.2.a | Introduce a topic and write to convey ideas and information about it including visual, tactual, or multimedia information as appropriate. |
| | EE.W.7.2.b | Provide facts, details, or other information related to the topic. |
| | EE.W.7.2.d | Select domain-specific vocabulary to use in writing about the topic. |

# Grade 8: Available Essential Elements and minimum expectation for each student's assessment

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.1** | **All students are assessed in this EE for C1.1** | |
| | EE.RI.8.5 | Locate the topic sentence and supporting details in a paragraph. |
| **ELA.C1.2** | **Choose at least three EEs in C1.2 (L, RL or RI) – EEs must be from at least two different strands.** | |
| | EE.RL.8.1 | Cite text to support inferences from stories and poems. |
| | EE.RL.8.2 | Recount an event related to the theme or central idea, including details about character and setting. |
| | EE.RL.8.4 | Determine connotative meanings of words and phrases in a text. |
| | EE.RI.8.1 | Cite text to support inferences from informational text. |
| | EE.RI.8.2 | Provide a summary of a familiar informational text. |
| | EE.RI.8.4 | Determine connotative meanings of words and phrases in a text. |
| | EE.RI.8.6 | Determine an author's purpose or point of view and identify examples from text to that describe or support it. |
| | EE.RI.8.8 | Determine the argument made by an author in an informational text. |
| | EE.L.8.5.a | Demonstrate understanding of the use of multiple meaning words. |
| **ELA.C1.3** | **Choose at least two EEs in C1.3, including at least one RL and one RI.** | |
| | EE.RL.8.3 | Identify which incidents in a story or drama lead to subsequent action. |
| | EE.RL.8.5 | Compare and contrast the structure of two or more texts. |
| | EE.RL.8.9 | Compare and contrast themes, patterns of events, or characters across two or more stories or dramas. |
| | EE.RI.8.3 | Recount events in the order they were presented in the text. |
| | EE.RI.8.9 | Identify where two different texts on the same topic differ in their interpretation of the details. |
| **ELA.C2.1** | **All students are assessed in all of these EEs through the writing assessment.** | |
| | EE.W.8.2.b | Write one or more facts or details related to the topic. |
| | EE.W.8.2.c | Write complete thoughts as appropriate. |
| | EE.W.8.2.d | Use domain specific vocabulary related to the topic. |
| | EE.W.8.2.f | Provide a closing. |
| | EE.W.8.2.a | Introduce a topic clearly and write to convey ideas and information about it including visual, tactual, or multimedia information as appropriate. |

# High School: Minimum expectation for each student's assessment in Grades 9-10 and Grades 11-12[1]

| Conceptual Area | EE | Description |
|---|---|---|
| **ELA.C1.2** | **Choose one EE in C1.2 (L, RL or RI).** | |
| | EE.RL.9-10.1 | Determine which citations demonstrate what the text says explicitly as well as inferences drawn from the text. |
| | EE.RL.9-10.2 | Recount events related to the theme or central idea, including details about character and setting. |
| | EE.RL.9-10.4 | Determine the meaning of words and phrases as they are used in a text, including idioms, analogies, and figures of speech. |
| | EE.RL.11-12.1 | Analyze a text to determine its meaning and cite textual evidence to support explicit and implicit understandings. |
| | EE.RL.11-12.2 | Recount the main events of the text which are related to the theme or central idea. |
| | EE.RL.11-12.4 | Determine how words or phrases in a text, including words with multiple meanings and figurative language, impacts the meaning. |
| | EE.RI.9-10.1 | Determine which citations demonstrate what the text says explicitly as well as inferentially. |
| | EE.RI.9-10.2 | Determine the central idea of the text and select details to support it. |
| | EE.RI.9-10.4 | Determine the meaning of words and phrases as they are used in text, including common idioms, analogies, and figures of speech. |
| | EE.RI.9-10.5 | Locate sentences that support an author's central idea or claim. |
| | EE.RI.9-10.8 | Determine how the specific claims support the argument made in an informational text. |
| | EE.RI.11-12.1 | Analyze a text to determine its meaning and cite textual evidence to support explicit and implicit understanding. |
| | EE.RI.11-12.2 | Determine the central idea of a text; recount the text. |
| | EE.RI.11-12.4 | Determine how words or phrases in a text, including words with multiple meanings and figurative language, impacts the meaning of the text. |
| | EE.RI.11-12.8 | Determine whether the claims and reasoning enhance the author's argument in an informational text. |
| | EE.RI.11-12.5 | Determine whether the structure of a text enhances an author's claim. |
| | EE.L.9-10.4.a | Use context to determine the meaning of unknown words. |
| | EE.L.9-10.5.b | Determine the intended meaning of multiple meaning words. |

---

[1] The high school blueprint provides coverage options for students in grades 9-12 to support the various testing requirements in different states in the consortium. Each state sets its own policy for which high school grade(s) are appropriate for DLM assessments.

| Conceptual Area | EE | Description |
|---|---|---|
| | EE.L.11-12.4.a | Use context to determine the meaning of unknown words. |
| **ELA.C1.3** | **Choose at least three EEs in C1.3 (RL or RI) – including at least one RL and one RI.** | |
| | EE.RL.9-10.3 | Determine how characters change or develop over the course of a text. |
| | EE.RL.9-10.5 | Identify where a text deviates from a chronological presentation of events. |
| | EE.RL.11-12.3 | Determine how characters, the setting or events change over the course of the story or drama. |
| | EE.RL.11-12.5 | Determine how the author's choice of where to end the story contributes to the meaning. |
| | EE.RI.9-10.3 | Determine logical connections between individuals, ideas or events in a text. |
| | EE.RI.11-12.3 | Determine how individuals, ideas, or events change over the course of the text. |
| | EE.RI.11-12.9 | Compare and contrast arguments made by two different texts on the same topic. |
| **ELA.C2.1** | **Choose at least two EEs in C2.1, including at least one W and one L** | |
| | EE.L.9-10.2.c | Spell most single-syllable words correctly and apply knowledge of word chunks in spelling longer words. |
| | EE.W.9-10.2.c | Use complete, simple sentences as appropriate. |
| | EE.W.9-10.2.d | Use domain specific vocabulary when writing claims related to a topic of study or text. |
| | EE.W.9-10.2.f | Provide a closing. |
| | EE.W.11-12.2.c | Use complete, simple sentences, as well as compound and other complex sentences as appropriate. |
| | EE.W.11-12.2.d | Use domain specific vocabulary when writing claims related to a topic of study or text. |
| | EE.W.11-12.2.f | Provide a closing or concluding statement. |
| | EE.L.11-12.2.b | Spell most single-syllable words correctly and apply knowledge of word chunks in spelling longer words. |
| **ELA.C2.2** | **All students are assessed in all the EEs identified for the appropriate grade level in both conceptual areas in Claim 2.** | |
| | EE.W.9-10.2.a | Introduce a topic clearly and use a clear organization to write about it including visual, tactual, or multimedia information as appropriate. |
| | EE.W.9-10.2.b | Develop the topic with facts or details. |
| | EE.W.11-12.2.a | Introduce a topic clearly and write an informative or explanatory text that conveys ideas, concepts, and information including visual, tactual, or multimedia information as appropriate. |
| | EE.W.11-12.2.b | Develop the topic with relevant facts, details, or quotes. |

Appendix S –     End-of-Course Blueprints

## Blueprint for ALGEBRA I

| Category | Code | Target | Point Range | Range Of Emphasis |
|---|---|---|---|---|
| Number & Quantity | HSN-RN.A | The Real Number System | 2-4 | 5-10% |
| | HSN-Q | Quantities | | |
| Algebra | HSA-SSE | Seeing Structure In Expressions | 14-21 | 35-53% |
| | HSA-APR | Arithmetic With Polynomials And Rational Expressions | | |
| | HSA-CED | Creating Equations | | |
| | HSA-REI | Reasoning With Equations And Inequalities | | |
| Functions | HSF-IF | Interpreting Functions | 11-20 | 28-50% |
| | HSF-BF | Building Functions | | |
| | HSF-LE | Linear, Quadratic And Exponential Models | | |
| Stats & Prob | HSS-ID | Interpreting Categorical And Quantitative Data | 3-6 | 8-15% |
| **Total** | | | 40 | 100% |
| **Performance Event:** Each year the performance event may align to any specific conceptual category or to a group of them. The Performance Event is worth 10 points. | | | | |

## Blueprint for ALGEBRA II

| Category | Code | Target | Point Range | Range Of Emphasis |
|---|---|---|---|---|
| Number & Quantity | HSN-CN | The Complex Number System | 0-4 | 0-10% |
| Algebra | HSA-SSE | Seeing Structure In Expressions | 16-22 | 40-55% |
| | HSA-APR | Arithmetic With Polynomials And Rational Expressions | | |
| | HSA-CED | Creating Equations | | |
| | HSA-REI | Reasoning With Equations And Inequalities | | |
| Functions | HSF-IF | Interpreting Functions | 18-24 | 45-60% |
| | HSF-BF | Building Functions | | |
| | HSF-LE | Linear, Quadratic And Exponential Models | | |
| Stats & Prob | HSS-ID | Interpreting Categorical And Quantitative Data | 0-6 | 0-15% |
| | HSS-MD | Using Probability To Make Decisions | | |
| **Total** | | | 40 | 100% |

## Blueprint for AMERICAN HISTORY

| Reporting Categories | Point Range | Range Of Emphasis |
|---|---|---|
| Government | 7-9 | 18%-23% |
| History | 14-18 | 35%-45% |
| Economics | 7-9 | 18%-23% |
| Geography | 7-9 | 18%-23% |
| **Total** | **40** | **100%** |

## Blueprint for BIOLOGY

| Content Strand | Point Range | Range Of Emphasis |
|---|---|---|
| Characteristics and Interactions of Living Organisms | 20-24 | 36%-44% |
| Changes in Ecosystems and Interactions of Organisms with their Environments | 12-14 | 22%-25% |
| Scientific Inquiry | 20 | 36% |
| **Total** | **55** | **100%** |

## Blueprint for ENGLISH I

| Claim | Category | Big Idea | Point Range | Range Of Emphasis |
|---|---|---|---|---|
| Reading | Claim 1a | Apply reading skills to demonstrate the ability to integrate key ideas and details, interpret and analyze the craft and structure of texts, and evaluate the knowledge and ideas found in literary texts | 15 | 33% |
| Reading | Claim 1b | Apply reading skills to demonstrate the ability to integrate key ideas and details, interpret and analyze the craft and structure of texts, and evaluate the knowledge and ideas found in informational text | 15 | 33% |
| Writing | Claim 2a | Demonstrate the ability to produce a variety of text types and purposes | 10 | 22% |
| Writing | Claim 2b | Demonstrate a command of the conventions of standard English, appropriate grade-level acquisition of vocabulary | 5 | 11% |
| **Total** | | | 45 | 100% |

## Blueprint for ENGLISH II

| Claim | Category | Big Idea | Point Range | Range Of Emphasis |
|---|---|---|---|---|
| Reading | Claim 1a | Apply reading skills to demonstrate the ability to integrate key ideas and details, interpret and analyze the craft and structure of texts, and evaluate the knowledge and ideas found in literary texts | 15 | 33% |
| Reading | Claim 1b | Apply reading skills to demonstrate the ability to integrate key ideas and details, interpret and analyze the craft and structure of texts, and evaluate the knowledge and ideas found in informational text | 15 | 33% |
| Writing | Claim 2a | Demonstrate the ability to produce a variety of text types and purposes | 10 | 22% |
| Writing | Claim 2b | Demonstrate a command of the conventions of standard English, appropriate grade-level acquisition of vocabulary | 5 | 11% |
| **Total** | | | 45 | 100% |

## Blueprint for GEOMETRY

| Category | Code | Target | Point Range | Range Of Emphasis |
|---|---|---|---|---|
| Geometry | HSF-CO | Congruence | 34-40 | 85-100% |
| | HSG-SRT | Similarity, Right Triangles And Trigonometry | | |
| | HSG-C | Circles | | |
| | HSG-GPE | Expressing Geometric Properties With Equations | | |
| | HSG-GMD | Geometric Measurement And Dimension | | |
| | HSG-MG | Linear, Quadratic And Exponential Models | | |
| Stats & Prob | HSS-CP | Conditional Probability And The Rules Of Probability | 0-6 | 0-15% |
| | HSS-MD | Using Probability To Make Decisions | | |
| **Total** | | | **40** | **100%** |

## Blueprint for GOVERNMENT

| Content Strand | Point Range | Range Of Emphasis |
|---|---|---|
| Principles of Constitutional Democracy | 18-22 | 45%-55% |
| Principles and Processes of Governance Systems | 18-22 | 45%-55% |
| **Total** | **40** | **100%** |

## Blueprint for PHYSICAL SCIENCE

| Content Strand | Point Range | Range Of Emphasis |
|---|---|---|
| Properties And Principles Of Matter And Energy | 25-30 | 55-66% |
| Properties And Principles Of Force And Motion | 15-20 | 33-44% |
| **Total** | **45** | **100%** |

Appendix T –     Assessment Testing Windows

# Assessment

The Assessment section provides professional services related to the Missouri Assessment Program (MAP) and the National Assessment of Educational Progress (NAEP).

The Assessment Section manages test development, on-going test maintenance, and oversees the test administration for four statewide, large-scale assessments. The MAP assessments test students' progress toward mastery of the Missouri Show-Me Standards.

## Assessment Calendar

| Assessment | Dates | Event |
|---|---|---|
| **Personal Finance** | June 9, 2014 to August 29, 2014 | Summer 2014 Window |
| **ACCESS for ELLs** | September 15, 2014 to September 26, 2014 | DTC Gathers ELL Roster and Tier Placement |
| **EOC** | **DUE** September 19, 2014 | Fall 2014 First Precode - Students Available in iTester 9/29 - Students May Begin Testing 10/6 |
| **ACCESS for ELLs** | September 29, 2014 to October 31, 2014 | Ordering ACCESS Test Materials |
| **ACCESS for ELLs** | September 29, 2014 to November 7, 2014 | Precoding for ACCESS for ELLs |
| **Personal Finance** | October 6, 2014 to January 23, 2015 | Fall 2014 Window |
| **EOC** | October 6, 2014 to January 23, 2015 | Fall 2014 Window |
| **EOC** | **DUE** November 7, 2014 | Fall 2014 Second Precode - Students Available in iTester 11/17 - Students May Begin Testing 11/24 |

| Assessment | Dates | Event |
|---|---|---|
| ACCESS for ELLs | December 12, 2014 | Test Materials Arrive In District |
| MAP-A | January 2, 2015 | Science - Transfer Student Participation Deadline |
| MAP-A | January 5, 2015 to January 30, 2015 | Science - Collection Period One |
| ACCESS for ELLs | January 12, 2015 to March 6, 2015 | ACCESS for ELLs TEST WINDOW |
| NAEP | January 26, 2015 to March 6, 2015 | 2015 Testing Window |
| EOC | **DUE** January 30, 2015 | Spring 2015 First Precode - Students Available in iTester 2/16/15 - Students May Begin Testing 2/23/15 |
| MAP-A | February 2, 2015 to February 27, 2015 | Science - Collection Period Two |
| ACCESS for ELLs | February 20, 2015 | Additional ACCESS for ELLs Materials Ordering Deadline |
| EOC | February 23, 2015 to May 22, 2015 | Spring 2015 Window |
| Personal Finance | February 23, 2015 to May 22, 2015 | Spring 2015 Window |
| ACCESS for ELLs | March 6, 2015 to March 20, 2015 | Districts Pack and Ship ACCESS for ELLs Materials |
| EOC | **DUE** March 6, 2015 | Spring 2015 Second Precode - Students Available in iTester 3/16/15 - Students May Begin Testing 3/23/15 |
| ACCESS for ELLs | March 20, 2015 | Deadline to Ship ACCESS for ELLs Materials |
| MAP-A | March 30, 2015 to May 22, 2015 | English language arts and Mathematics - Dynamic Learning Maps (Year-Ends) |

| Assessment | Dates | Event |
|---|---|---|
| **NAEP** | March 30, 2015 to May 29, 2015 | TIMSS Testing Window |
| **Grade-Level** | March 30, 2015 to May 22, 2015 | MAP Grade-Level Assessments Window |
| **EOC** | **DUE** April 3, 2015 | Spring 2015 Third Precode - Students Available in iTester 4/20/15 - Students May Begin Testing 4/27/15 |
| **ACCESS for ELLs** | May 20, 2015 to May 22, 2015 | Reports Arrive In District |

| **Grade-Level** | **End-of-Course** | **ACCESS** | **Personal Finance** | **NAEP** | **MAP-A** |
|---|---|---|---|---|---|

- **End-of-Course**
- **2014-2015 LEA Guide To The Missouri Assessment Program**
- **iTester Administration**
- **Missouri Learning Standards**

# Peer Review Status

All states' assessment systems are submitted to the United States Department of Education for Peer Review to document that they meet the requirements of the No Child Left Behind Act of 2001. The current Peer Review status of all components of the Missouri Assessment Program is as follows:

| Assessment | Status |
|---|---|
| Grades 3 - 8 English Language Arts and Mathematics | Full Approval |
| Grades 5 and 8 Science | Full Approval |
| Algebra I, Biology, and English II | Full Approval |

Appendix U –     Sample Student Report

## Sample Reports

### Individual Student Report

The Individual Student Report provides information about performance on the End-of-Course Assessment, describing the results in terms of four levels of achievement in a content area. It is used for measuring and reflecting an individual's student mastery toward post-secondary readiness for a content area. It is used in instructional planning as a point of reference during a parent/teacher conference and for permanent record keeping. Other sources of information should be used along with this report when determining the student's areas of strength or need.

Achievement-level scores describe what students can do in terms of the Course-Level Expectations for the content and skills assessed by the End-of-Course Assessment. Students in the Proficient or Advanced levels have met the standard. Students in the Below Basic or Basic levels need to work on the skills described for their level on pages 8–15, as well as on skills in the next higher level.

The next page includes a sample of the Individual Student Report. The following areas on the sample have been identified to better explain the results that are being reported:

[A] The heading of the Individual Student Report includes the content area for the results being presented. A separate report is produced for each content area tested.

[B] The Student Information section contains the biographic data for the individual student taking the assessment. Identifying information for the MOSIS ID, gender, group, building, district, and test period are listed.

[C] The individual student's results are presented numerically as a three-digit scale score with the standard error (SE). An accompanying bar graph illustrates the achievement level obtained by the student. Achievement levels (whether Advanced, Proficient, Basic, or Below Basic) are based on the scale score ranges listed beneath the Achievement Scores heading in the table.

[D] The mean scale scores for the student's building and district are displayed in the two rows below the student's individual results. The mean scale score, with an associated SE, and the bar graph provide a way to view the individual's results in contrast to the group's results for the content area during the same test period.

[E] The narrative describes the student performance characteristics corresponding to the level of achievement obtained. The text is specific to the content area tested. At the bottom of the page is the URL, which provides additional information for all of the achievement levels for the content area.

# Missouri
### DEPARTMENT OF ELEMENTARY & SECONDARY
## EDUCATION™

**A** **End-of-Course Assessment**

**English II**

**B** **Individual Student Report for:**

**Jane E Dow**

| | |
|---|---|
| MOSIS ID: | 1536879236 |
| Gender: | F |
| Building: | Washington HS |
| Building Code: | 9999 |
| District: | Jefferson SD |
| District Code: | 999999 |
| Test Period: | Summer 2014 |

**C** **Achievement Scores**

| Below Basic 100-179 | Basic 180-199 | Proficient 200-224 | Advanced 225-250 |
|---|---|---|---|

Jane E Dow — **218** (SE ±5)

Washington HS — **205** (SE ±4)

Jefferson SD — **190** (SE ±1)

**D**

**E** **Jane E Dow Achievement Level: Proficient**

Students performing at the Proficient level on the Missouri English II End-of-Course Assessment demonstrate an understanding of the skills and processes identified in the Course Level Expectations for English II. They demonstrate these skills in reading processes and in responding to both fiction and nonfiction texts. In addition to understanding and applying the skills at the Basic level, students scoring at the Proficient level use a range of strategies to comprehend and interpret a variety of texts, demonstrate an understanding of literary forms, and apply strategies for accessing and summarizing information. They correctly apply the rules and conventions of Standard English.

**About Achievement Levels**

| Below Basic 100-179 | Basic 180-199 | Proficient 200-224 | Advanced 225-250 |
|---|---|---|---|
| Students demonstrate little understanding of the skills and processes identified in the Course Level Expectations for English II. | Students demonstrate an incomplete understanding of the skills and processes identified in the Course Level Expectations for English II. | Students demonstrate an understanding of the skills and processes identified in the Course Level Expectations for English II. | Students demonstrate a thorough understanding of the skills and processes identified in the Course Level Expectations for English II. |

For more information about achievement levels, please visit the following web site:
**http://dese.mo.gov/college-career-readiness/assessment/end-course/general-resources**

17